



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification⁶ : C07K 14/00	A2	(11) International Publication Number: WO 98/25959 (43) International Publication Date: 18 June 1998 (18.06.98)
(21) International Application Number: PCT/US97/22787 (22) International Filing Date: 11 December 1997 (11.12.97) (30) Priority Data: 60/032,757 11 December 1996 (11.12.96) US (71) Applicant: CHIRON CORPORATION [US/US]; 4560 Horton Street, Emeryville, CA 94608 (US). (72) Inventors: ESCOBEDO, Jaime; 1470 Livorna Road, Alamo, CA 94507 (US). HU, Quianjin; Chiron Corporation, 4560 Horton Street, Emeryville, CA 94608 (US). GARCIA, Pablo; 882 Chenery Street, San Francisco, CA 94131 (US). WILLIAMS, Lewis, T.; 3 Miraflores Lane, Tibouron, CA 94920 (US). KOTHAKOTA, Srinivas; Chiron Corporation, 4560 Horton Street, Emeryville, CA 94608 (US). (74) Agents: POTTER, Jane, E., R.; Chiron Corporation, Intellectual Property - R440, P.O. Box 8097, Emeryville, CA 94662-8097 (US) et al.		(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG). Published <i>Without international search report and to be republished upon receipt of that report.</i>
(54) Title: SECRETED HUMAN PROTEINS (57) Abstract Secreted proteins can be identified using a method which exploits the ability of microsomes to modify proteins post-translationally. Nineteen human secreted proteins and full-length cDNA sequences encoding the proteins have been identified using this method. The proteins and cDNA sequences can be used, <i>inter alia</i> , for targeting other proteins to the membrane or extracellular milieu.		

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakhstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

SECRETED HUMAN PROTEINS

5 This application claims the benefit of copending provisional application
Serial No. 60/032,757, filed December 11, 1996, which is incorporated herein by
reference.

TECHNICAL AREA OF THE INVENTION

10 The invention relates to the area of proteins. More particularly, the
invention relates to human secreted proteins.

BACKGROUND OF THE INVENTION

15 Secreted proteins include such important proteins as growth factors,
cytokines and their receptors, extracellular matrix proteins, and proteases.
Nucleotide sequences encoding these proteins can be used to detect disease states in
which such proteins are implicated and to develop therapeutics for such diseases.
Thus, there is a need in the art for methods of identifying secreted proteins and the
nucleotide sequences which encode them.

SUMMARY OF THE INVENTION

20 It is an object of the invention to provide an isolated and purified human
protein.

It is yet another object of the invention to provide a fusion protein.

It is still another object of the invention to provide a preparation of antibodies.

It is even another object of the invention to provide an isolated and purified subgenomic polynucleotide.

5 It is yet another object of the invention to provide an isolated gene.

It is a further object of the invention to provide a DNA construct for expressing all or a portion of a human protein.

It is still another object of the invention to provide a host cell comprising a DNA construct.

10 It is another object of the invention to provide a homologously recombinant cell.

It is even another object of the invention to provide a method of producing a human protein.

15 It is another object of the invention to provide a method of identifying a secreted polypeptide which is modified by rough microsomes.

These and other objects of the invention are provided by one or more of the embodiments described below.

One embodiment of the invention provides an isolated and purified human protein. The isolated and purified human protein has an amino acid sequence
20 selected from the group consisting of the amino acid sequences shown in SEQ ID Nos:20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, and 38.

Another embodiment of the invention provides an isolated and purified human protein having an amino acid sequence which is at least 85% identical to an amino acid sequence selected from the group consisting of the amino acid
25 sequences shown in SEQ ID Nos:20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, and 38.

Still another embodiment of the invention provides a polypeptide comprising at least 6 contiguous amino acids of an amino acid sequence selected from the group consisting of the amino acid sequences shown in SEQ ID Nos:20, 21, 22, 23,
30 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, and 38.

Even another embodiment of the invention provides a fusion protein. The fusion protein comprises a first protein segment and a second protein segment fused together by means of a peptide bond. The first protein segment consists of at least 6 contiguous amino acids selected from the group consisting of the amino acid sequences shown in SEQ ID Nos:20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, and 38.

Yet another embodiment of the invention provides a preparation of antibodies. The antibodies specifically bind to a human protein having an amino acid sequence selected from the group consisting of the amino acid sequences shown in SEQ ID Nos:20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, and 38.

Even another embodiment of the invention provides an isolated and purified subgenomic polynucleotide. The isolated and purified subgenomic polynucleotide has a nucleotide sequence selected from the group consisting of the nucleotide sequences shown in SEQ ID NOs:1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, and 19.

Yet another embodiment of the invention provides an isolated and purified subgenomic polynucleotide consisting of at least 10 contiguous nucleotides selected from the group consisting of the nucleotide sequences shown in SEQ ID NOs:1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, and 19.

Still another embodiment of the invention provides an isolated gene. The isolated gene corresponds to a cDNA sequence selected from the group consisting of the nucleotide sequences shown in SEQ ID NOs:1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, and 19.

Another embodiment of the invention provides a DNA construct for expressing all or a portion of a human protein. The DNA construct comprises a promoter and a polynucleotide segment. The polynucleotide segment encodes at least 6 contiguous amino acids of a human protein having an amino acid sequence selected from the group consisting of the amino acid sequences shown in SEQ ID Nos:20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, and 38.

The polynucleotide segment is located downstream from the promoter.

Transcription of the polynucleotide segment initiates at the promoter.

Even another embodiment of the invention provides a host cell comprising a DNA construct. The DNA construct comprises a promoter and a polynucleotide segment. The polynucleotide segment encodes at least 6 contiguous amino acids of a human protein having an amino acid sequence selected from the group consisting of the amino acid sequences shown in SEQ ID Nos:20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, and 38. The polynucleotide segment is located downstream from the promoter. Transcription of the polynucleotide segment initiates at the promoter.

Still another embodiment of the invention provides a homologously recombinant cell having incorporated therein a new transcription initiation unit. The transcription initiation unit comprises in 5' to 3' order an exogenous regulatory sequence, an exogenous exon, and a splice donor site. The transcription initiation unit is located upstream to a coding sequence of a gene. The gene comprises a nucleotide sequence selected from the group consisting of the nucleotide sequences shown in SEQ ID NOs:1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, and 19. The exogenous regulatory sequence controls transcription of the coding sequence of the gene.

Yet another embodiment of the invention provides a method of producing a human protein. A culture of a cell is grown. The cell comprises a DNA construct. The DNA construct comprises a promoter and a polynucleotide segment. The polynucleotide segment encodes at least 6 contiguous amino acids of a human protein having an amino acid sequence selected from the group consisting of the amino acid sequences shown in SEQ ID Nos:20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, and 38. The polynucleotide segment is located downstream from the promoter. Transcription of the polynucleotide segment initiates at the promoter. The protein is purified from the culture.

Even another embodiment of the invention provides a method of producing a human protein. A culture of a cell is grown. The cell comprises a new transcription initiation unit. The transcription initiation unit comprises in 5' to 3'

order an exogenous regulatory sequence, an exogenous exon, and a splice donor site. The transcription initiation unit is located upstream to a coding sequence of a gene. The gene comprises a nucleotide sequence selected from the group consisting of the nucleotide sequences shown in SEQ ID NOs:1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, and 19. The exogenous regulatory sequence controls transcription of the coding sequence of the gene. The protein is purified from the culture.

Another embodiment of the invention provides a method of identifying a secreted polypeptide which is modified by rough microsomes. A population of cDNA molecules is transcribed *in vitro* whereby a population of cRNA molecules is formed. A first portion of the population of cRNA molecules is translated *in vitro* in the absence of rough microsomes whereby a first population of polypeptides is formed. A second portion of the population of cRNA molecules is translated *in vitro* in the presence of rough microsomes whereby a second population of polypeptides is formed. The first population of polypeptides is compared with the second population of polypeptides. Polypeptide members of the second population which have been modified by the rough microsomes are detected.

The present invention thus provides the art with a method for identifying secreted proteins or polypeptides, the amino acid sequences of nineteen novel human secreted proteins, and the nucleotide sequences which encode these proteins. The invention can be used to, *inter alia*, to produce secreted proteins for therapeutic and diagnostic purposes.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

The inventors have discovered a method for identifying secreted proteins or polypeptides. Secreted proteins or polypeptides include soluble proteins which can be transported across a membrane, such as a cell membrane, nuclear membrane, or membrane of the endoplasmic reticulum, as well as proteins which can be partially secreted from a cell, such as membrane-bound receptors.

Secreted proteins can contain a signal (or secretion leader) sequence, located at the N-terminus and including at least several hydrophobic amino acids,

such as phenylalanine, methionine, leucine, valine, or tryptophan. Non-hydrophobic amino acids can also be included in the signal sequence. Signal sequences are described in von Heijne, *J. Mol. Biol.* 184:99-105 (1985) and Kaiser and Botstein, *Mol. Cell. Biol.* 6:2382-2391 (1986). Secreted proteins can also be glycosylated by post-translational modification. The presence of a signal sequence or the presence of glycosylation or both indicate that a particular protein is a secreted protein.

In order to identify secreted proteins or polypeptides, the method of the invention exploits properties of microsomes, which are the closed vesicles that result from fragmentation of endoplasmic reticulum. Microsomes can be rough or smooth, depending on whether the endoplasmic reticulum from which they were derived is studded with ribosomes. Microsomes, particularly rough microsomes, have the ability to perform post-translational modifications, such as glycosylation and cleavage of signal sequences from proteins or polypeptides.

To identify secreted proteins, a population of complementary DNA (cDNA) molecules is transcribed *in vitro* to synthesize a population of complementary RNA (cRNA) molecules. The cDNA molecules can be synthesized by reverse transcription of mRNA molecules isolated from a particular cell or tissue type or organism using, for example, a commercially available reverse transcriptase enzyme. Alternatively, the reverse transcription reaction to form cDNA molecules can be conducted on total RNA, without a preliminary purification of mRNA.

Any organism, such as a bacterium, plant, invertebrate, or vertebrate organism, can be used as a source of RNA. Particularly preferred sources of RNA are mammals, most preferably humans. Tissues, such as liver, brain, kidney, spleen, pancreas, or muscle, can be used as a source of RNA. Individual cell types, either primary cells or members of established cell lines, such as HeLa, CHO, PC12, P19, BHK, COS, or HepG2, are suitable sources of RNA. Tissues or primary cells isolated from organisms at a particular stage in development can be used as RNA sources. Stem cells, such as hematopoietic, neuronal, and embryonic stem cells, can also be used as a source of RNA.

Total RNA or mRNA can be isolated using methods known in the art. Such methods are described, *inter alia*, in Sambrook *et al.*, MOLECULAR CLONING, A

LABORATORY MANUAL (2d ed., Cold Spring Harbor Press, N.Y., 1989), and Ausubel *et al.*, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY (Greene Publishing Associates and John Wiley & Sons, N.Y., 1994). Techniques for RNA isolation can be tailored for a particular organism or cell type, as is known in the art.

5 Complementary DNA can optionally be obtained from a cDNA library. The cDNA library can be derived from the genome of any organism of interest, particularly a mammal or a human. Tissue- or cell type-specific cDNA libraries can also be used as a source of cDNA.

10 Transcription of cDNA molecules *in vitro* to form cRNA molecules can be carried out using any methods known in the art. These methods include, for example, placing cDNA into a cloning vector containing a promoter, such as an SP6, T7, or T3 polymerase promoter, and transcribing the cDNA using the appropriate polymerase. A variety of commercial kits are available for this purpose.

15 A first portion of the population of cRNA molecules can be translated *in vitro*, in the absence of rough microsomes, to form a first population of polypeptides which have not been post-translationally modified. A second portion of the population of cRNA molecules can be translated *in vitro* in the presence of rough microsomes. Under the conditions of the *in vitro* translation reaction, rough microsomes can cleave signal sequences from those polypeptides which comprise
20 such sequences. Under the same conditions, rough microsomes can also glycosylate those polypeptides which contain glycosylation sites.

25 Methods of *in vitro* translation are those which are known in the art, such as translation in a reticulocyte lysate system, particularly a rabbit reticulocyte lysate. Reticulocyte lysate systems can be assembled in the laboratory or purchased commercially in kit form.

30 Microsomes can be prepared by disruption of tissues or cells by homogenization, as is known in the art. If desired, rough and smooth microsomes can be separated using well-known techniques, such as sucrose density gradient sedimentation. Microsomes are also available commercially, for example, such as the canine pancreatic microsomes available from Promega Corp., Madison, WI.

The first population of polypeptides can then be compared with the second population of polypeptides. This comparison can be by means of, for example, one- or two-dimensional polyacrylamide gel electrophoresis, as is known in the art. Polypeptides separated in the gels can be detected by any means known in the art, such as staining with copper, silver, Coomassie Brilliant Blue, amido black, fast green FCF, Ponceau S, or a chromophoric label. Separated proteins can also be visualized using radioactive, chemiluminescent, fluorescent, or enzymatic tags incorporated into the proteins before separation.

The gels can be dried or the proteins can be transferred to membranes, such as polyvinylidene difluoride membranes. Either the gels or membranes themselves or photographs of the gels or membranes can be compared by eye. Alternatively, the gels or membranes can be scanned, for example, with a densitometer and analyzed with the aid of a computer.

Polypeptide members of the second population of polypeptides, which have been modified by the rough microsomes, can be detected by any means available in the art. For example, a shift in the position of a polypeptide band can be observed, indicating an increase in molecular weight of a member of the second population compared with the corresponding polypeptide member of the first population. Such an increase in molecular weight indicates that the polypeptide member of the second population was glycosylated by the rough microsomes.

A shift in the position of a polypeptide band indicating a decrease in molecular weight of a member of the second population compared with the corresponding polypeptide member of the first population can also be observed. This decrease in molecular weight indicates that the polypeptide member of the second population contained a signal sequence which was cleaved by the rough microsomes.

Polypeptides which are modified by the rough microsomes are identified as secreted polypeptides. Optionally, quantities of cDNA molecules which encode secreted polypeptides can be obtained. Molecules of cDNA which encode polypeptides which are post-translationally modified by the rough microsomes can be placed into suitable vectors using standard recombinant DNA techniques and

used to transform host cells. Many vectors are available for this purpose, such as retroviral or adenoviral vectors and bacteriophage, as described below.

Vectors comprising cDNA which encode secreted polypeptides can be introduced into host cells using techniques available in the art. These techniques include, but are not limited to, transferrin-polycation-mediated DNA transfer, transfection with naked or encapsulated nucleic acids, liposome-mediated cellular fusion, intracellular transportation of DNA-coated latex beads, protoplast fusion, viral infection, electroporation, and calcium phosphate-mediated transfection.

The host cells can be any host cells which are capable of propagating cDNA molecules. A variety of host cells, for example immortalized cell lines such as HeLa, CHO, or HEK, are available for this purpose.

Transformed host cells can be diluted serially and cultured to form individual colonies. Methods of culturing host cells and the media suitable for each host cell type are well known in the art. Preferably, each colony originates from a single transformed host cell. Separate preparations of cDNA from each colony can be prepared, as described above, and transcribed *in vitro* to form cRNA. The cRNA can be transcribed to form secreted polypeptides, which can be purified as is known in the art. If the preparation of secreted polypeptides from a colony contains more than one species of polypeptide, the steps described above can be repeated until a colony is obtained which contains cDNA encoding only a single species of polypeptide.

Complementary DNA molecules which encode secreted proteins can be sequenced using standard nucleotide sequencing techniques. The sequence of each cDNA molecule can be compared with known sequences in a database to determine whether the clone encodes a known or a novel secreted protein.

The inventors have used the method of the invention to identify nineteen novel human secreted proteins. Amino acid sequences for these nineteen human secreted proteins are disclosed in SEQ ID Nos:20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, and 38. Nucleotide sequences which encode the proteins are disclosed in SEQ ID NOs:1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, and 19, respectively.

Clones containing the cDNAs of the secreted proteins were deposited on December 11, 1997, with the ATCC. Individual bacterial cells (*E. coli*) in this composite deposit contain one or more of the polynucleotides encoding the secreted proteins of the invention and can be retrieved using an oligonucleotide probe designed from the sequence for that particular polynucleotide, as provided herein. Each polynucleotide can be removed from the vector by performing an EcoRI/NotI digestion (5' site, EcoRI; 3' site, NotI). The deposit submitted to the ATCC has been designated SECP120997. The nucleotide sequences of these deposits and the amino acid sequences they encode are controlling in the event of a discrepancy between the amino acid and nucleotide sequences disclosed herein and those contained in the deposits.

A purified and isolated subgenomic polynucleotide of the present invention comprises at least 10, 12, 15, 18, 20, 25, 30, 35, 40, 45, or 50 contiguous nucleotides selected from the nucleotide sequences shown in SEQ ID NOs:1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, and 19. The isolated and purified subgenomic polynucleotides can comprise an entire nucleotide sequence selected from the nucleotide sequences shown in SEQ ID NOs:1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, and 19.

Subgenomic polynucleotides contain less than a whole chromosome and are preferably intron-free. Polynucleotides of the invention can be isolated and purified free from other nucleotide sequences by standard nucleic acid purification techniques, using restriction enzymes and probes to isolate fragments comprising the coding sequences.

Isolated genes corresponding to the cDNA sequences disclosed herein are also provided. Known methods can be used to isolate the corresponding genes using the provided cDNA sequences. These methods include preparation of probes or primers from the nucleotide sequences shown in SEQ ID NOs:1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, and 19 for use in identifying or amplifying the genes from human genomic libraries or other sources of human genomic DNA.

The coding sequences shown in SEQ ID NOs:1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, and 19 can be made using reverse transcriptase with

human mRNA as a template. Amplification by PCR can also be used to obtain the polynucleotides, using either genomic DNA or cDNA as a template. Polynucleotide molecules of the invention can also be made using the techniques of synthetic chemistry given the sequences disclosed herein. The degeneracy of the genetic code permits alternate nucleotide sequences which will encode the amino acid sequences shown in SEQ ID Nos:20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, and 38 to be synthesized. All such nucleotide sequences are within the scope of the present invention.

Polynucleotide molecules of the invention can be propagated in vectors and cell lines as is known in the art. Polynucleotide molecules can be on linear or circular molecules. They can be on autonomously replicating molecules or on molecules without replication sequences. For propagation, polynucleotides of the invention can be introduced into suitable host cells using any techniques available in the art, as described above.

Subgenomic polynucleotides of the invention can be used to propagate additional copies of the polynucleotides or to express protein, polypeptides, or fusion proteins. The subgenomic polynucleotides disclosed herein can also be used, for example, as biomarkers for tissues or chromosomes, as molecular weight markers for DNA gels, to elicit immune responses, such as the formation of antibodies against single- or double-stranded DNA, and in DNA-ligand interaction assays, to detect proteins or other molecules which interact with the nucleotide sequences.

Disease states may be associated with alterations in the expression of genes which encode proteins of the invention. Polynucleotide sequences disclosed herein can also be used to determine the involvement of any of these sequences in disease states. For example, a gene in a diseased cell can be sequenced and compared with a wild-type coding sequence of the invention. Alternatively, nucleotide probes can be constructed and used to detect normal or altered (mutant) forms of mRNA in a diseased cell. Subgenomic polynucleotides of the invention can also be used to design diagnostic tests and therapeutic compositions for diseases which may be associated with altered expression of these genes.

The present invention provides both full-length and mature forms of the disclosed proteins. Full-length forms of the proteins have the amino acid sequences shown in SEQ ID Nos:20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, and 38. The full-length forms of a protein can be processed enzymatically to remove a signal sequence, resulting in a mature form of the protein. Signal sequences can be identified by examination of the amino acid sequences disclosed herein and comparison with amino acid sequences of known signal sequences (see, *e.g.*, von Heijne, 1985; Kaiser & Botstein, 1986). Similarly, transmembrane domains can be identified by examination of the amino acid sequences disclosed herein. A transmembrane domain typically contains a long stretch of 15-30 hydrophobic amino acids.

Other domains with predicted functions can also be identified. For example, the protein having the amino acid sequence shown in SEQ ID NO:23 comprises a Kunitz type serine protease inhibitor domain spanning amino acids 68 to 122 of SEQ ID NO:23. The protein having the amino acid sequence shown in SEQ ID NO:20 contains a zinc-finger motif.

Allelic variants of the disclosed subgenomic polynucleotides can occur and encode proteins which are identical, homologous, or substantially related to amino acid sequences disclosed herein (see below).

Allelic variants of subgenomic polynucleotides of the invention can be identified by hybridization of putative allelic variants with nucleotide sequences disclosed herein under stringent conditions. For example, by using the following wash conditions--2 x SCC, 0.1% SDS, room temperature twice, 30 minutes each; then 2 x SCC, 0.1% SDS, 50 °C. once, 30 minutes; then 2 x SCC, room temperature twice, 10 minutes each--allelic variants can be identified which contain at most about 25-30% basepair mismatches. More preferably, allelic variants contain 15-25% basepair mismatches, even more preferably 5-15% basepair mismatches.

Protein variants of secreted proteins of the invention are also included. Amino acids which are not involved in regions which determine biological activity can be deleted or modified without affecting biological function. Preferably, protein

variants of the invention have amino acid sequences which are at least 85%, 90%, or 95% identical to the amino acid sequences disclosed herein and have similar biological properties (see below). More preferably, the molecules are 98% identical. Modifications of interest in the protein sequences can include the alteration, substitution, replacement, insertion or deletion of a selected amino acid residue. Proteins or derivatives can be either glycosylated or unglycosylated. Techniques for making such modifications are well known to those skilled in the art (*see, e.g.*, U.S. 4,518,584). Alternatively, variants of proteins disclosed herein can be constructed using techniques of synthetic chemistry or using recombinant DNA methods.

Preferably, amino acid changes in variants or derivatives of proteins of the invention are conservative amino acid changes, *i.e.*, substitutions of similarly charged or uncharged amino acids. A conservative amino acid change involves substitution of one amino acid for another amino acid of a family of amino acids which are structurally related in their side chains. Naturally occurring amino acids are generally divided into four families: acidic (aspartate, glutamate), basic (lysine, arginine, histidine), non-polar (alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), and uncharged polar (glycine, asparagine, glutamine, cystine, serine, threonine, tyrosine) amino acids. Phenylalanine, tryptophan, and tyrosine are sometimes classified as aromatic amino acids. It is reasonable to expect that an isolated replacement of a leucine with an isoleucine or valine, an aspartate with a glutamate, a threonine with a serine, or a similar replacement of an amino acid with a structurally related amino acid will not have a major effect on the binding properties of the resulting molecule, especially if the replacement does not involve an amino acid at a binding site involved in an interaction of the protein. Non-naturally occurring amino acids can also be used to form protein variants of the invention.

Whether an amino acid change results in a functional protein or polypeptide can readily be determined by assaying biological properties of the disclosed proteins or polypeptides, as described below. Species homologs of human subgenomic polynucleotides and proteins of the invention can also be identified by making

suitable probes or primers and screening cDNA expression libraries from other species, such as mice, monkeys, yeast, or bacteria.

In the case of proteins which are membrane-bound, such as cell surface receptor proteins, soluble forms of the proteins can be obtained by deleting the nucleotide sequences which encode part or all of the intracellular and transmembrane domains of the protein and expressing a fully secreted form of the protein in a host cell. Techniques for identifying intracellular and transmembrane domains, such as homology searches, can be used to identify such domains in proteins of the invention using amino acid and nucleotide sequences disclosed herein.

Polypeptides consisting of less than full-length proteins of the present invention are also provided. Polypeptides of the invention can be linear or can be cyclized, for example, as described in Saragovi *et al.*, 1992, *Bio/Technology* 10, 773-778 and McDowell *et al.*, 1992, *J. Amer. Chem. Soc.* 114, 9245-9253.

Polypeptides can be used, for example, as immunogens, diagnostic aids, or therapeutics, and to create fusion proteins, as described below.

Polypeptide molecules consisting of less than the entire amino acid sequences shown in SEQ ID Nos:20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, and 38 are also provided. Such polypeptides comprise at least 6, 8, 10, 12, 15, 18, or 20 contiguous amino acids of an amino acid sequence shown in SEQ ID Nos:20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, and 38. Polypeptide molecules of the invention can also possess minor amino acid alterations which do not substantially affect the ability of the polypeptides to interact with specific molecules, such as antibodies.

Derivatives of the polypeptides, such as glycosylated forms, aggregative conjugates with other molecules, and covalent conjugates with unrelated chemical moieties, are also provided. Derivatives also include allelic variants, species variants, and muteins. Covalent derivatives are prepared by linkage of functionalities to groups which are found in the amino acid chain or at the N- or C-terminal residue by means known in the art. Truncations or deletions of regions which do not affect biological function are also encompassed. Truncated or deleted

polypeptides can be prepared synthetically or recombinantly, or by proteolytic digestion of purified or partially purified secreted proteins of the invention.

5 Fusion proteins comprising at least 6, 8, 10, 12, 15, 18, or 20 contiguous amino acids of the disclosed proteins can also be constructed. Human fusion proteins are useful, *inter alia*, for generating antibodies against amino acid sequences and for use in various assay systems. For example, fusion proteins can be used to identify proteins which interact with secreted proteins of the invention and influence their function. Physical methods, such as protein affinity chromatography, or library-based assays for protein-protein interactions, such as the
10 yeast two-hybrid or phage display systems, can be used for this purpose. Such methods are well known in the art and can also be used as drug screens. Fusion proteins can also be used to target molecules to a specific location in a cell or to cause a molecule to be secreted or to be anchored in a cellular membrane.

15 Fusion proteins of the invention comprise two protein segments which are fused together with a peptide bond. The first protein segment comprises at least 6, 8, 10, 12, 15, 18, or 20 contiguous amino acids selected from an amino acid sequence shown in SEQ ID Nos:20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, and 38. The first protein segment can also be a full-length protein (comprising a signal sequence) or a mature protein (lacking a signal
20 sequence). The second protein segment can be a full-length protein or a protein fragment. The second protein or protein fragment can be labeled with a detectable marker, such as a radioactive, chemiluminescent, biotinylated, or fluorescent tag, or can be an enzyme which will generate a detectable product. Enzymes suitable for this purpose, such as β -galactosidase, are well known in the art.

25 Techniques for making fusion proteins, either recombinantly or by covalently linking two protein segments, are well known in the art. Fusion proteins comprising amino acid sequences of the invention can also be constructed, for example, using standard recombinant DNA methods to make a DNA construct which comprises contiguous nucleotides selected from SEQ ID NOs:1, 2, 3, 4, 5, 6,
30 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, and 19 and encoding the desired amino

acids in proper reading frame with nucleotides encoding the second protein segment.

Proteins or polypeptides of the invention can be purified free from other components with which they are normally associated in a cell, such as carbohydrates, lipids, subcellular organelles, or other proteins. An isolated protein or polypeptide is at least 90% pure. Preferably, the preparations are 95% or 99% pure. The purity of a preparation can be assessed, for example, by examining electrophoretograms of protein or polypeptide preparations at several pH values and at several polyacrylamide concentrations, as is known in the art.

Standard biochemical methods can be used to isolate proteins of the invention from tissues which express the proteins or to isolate proteins, polypeptides, or fusion proteins from recombinant host cells into which a DNA construct has been introduced. Methods of protein purification, such as size exclusion chromatography, ammonium sulfate fractionation, ion exchange chromatography, affinity chromatography, crystallization, electrofocusing, or preparative gel electrophoresis, are well known and widely used in the art.

Alternatively, proteins, fusion proteins, or polypeptides of the invention can be produced by recombinant DNA methods or by synthetic chemical methods.

~~Synthetic chemistry methods, such as solid phase peptide synthesis, can be used to~~ synthesize proteins, fusion proteins, or polypeptides. For production of recombinant proteins, fusion proteins, or polypeptides, coding sequences selected from the nucleotide sequences shown in SEQ ID NOs:1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, and 19 can be expressed in prokaryotic or eukaryotic host cells using expression systems known in the art. These expression systems include bacterial, yeast, insect, and mammalian cells (see below).

The resulting expressed protein can then be purified from the culture medium or from extracts of the cultured cells using purification procedures known in the art. For example, for proteins fully secreted into the culture medium, cell-free medium can be diluted with sodium acetate and contacted with a cation exchange resin, followed by hydrophobic interaction chromatography. Using this method, the desired protein, fusion protein, or polypeptide is typically greater than 95% pure.

Further purification can be undertaken, using, for example, any of the techniques listed above. Proteins, fusion proteins, or polypeptides can also be tagged with an epitope, such as a "Flag" epitope (Kodak), and purified using an antibody which specifically binds to that epitope.

5 It may be necessary to modify a protein produced in yeast or bacteria, for example by phosphorylation or glycosylation of the appropriate sites, in order to obtain a functional protein. Such covalent attachments can be made using known chemical or enzymatic methods.

10 Proteins or polypeptides of the invention can also be expressed in cultured cells in a form which will facilitate purification. For example, a secreted protein or polypeptide can be expressed as a fusion protein comprising, for example, maltose binding protein, glutathione-S-transferase, or thioredoxin, and purified using a commercially available kit. Kits for expression and purification of such fusion proteins are available from companies such as New England BioLabs, Pharmacia,
15 and Invitrogen.

The coding sequences disclosed herein can also be used to construct transgenic animals, such as cows, goats, pigs, or sheep. Female transgenic animals can then produce proteins, polypeptides, or fusion proteins of the invention in their milk. Methods for constructing such animals are known and widely used in the art.

20 Isolated proteins, polypeptides, or fusion proteins of the invention can be used to obtain a preparation of antibodies which specifically bind to epitopes comprising amino acid sequences of the invention. Antibodies of the invention can be used, for example, to detect proteins, polypeptides, or fusion proteins of the invention which are secreted into culture medium or to identify tissues or cells
25 which express these molecules. The antibodies can be polyclonal or monoclonal or can be single chain antibodies. Techniques for raising polyclonal and monoclonal antibodies and for constructing single chain antibodies are well known in the art.

30 Antibodies of the invention bind specifically to epitopes comprising amino acid sequences of the invention, preferably to epitopes not present on other proteins. Typically a minimum number of contiguous amino acids to encode an epitope is 6, 8, or 10. However, more amino acids can be part of an epitope, for

example, at least 15, 25, or 50, especially to form epitopes which involve non-contiguous residues. Specific binding antibodies do not detect other proteins on Western blots of proteins or in immunocytochemical assays. Specific binding antibodies provide a signal at least ten-fold lower than the signal provided with epitopes which do not comprise amino acid sequences of the invention. Antibodies which bind specifically to secreted proteins of the invention include those that bind to mature or full-length proteins, to polypeptides or degradation products, to fusion proteins, or to protein variants. In a preferred embodiment of the invention, the antibodies immunoprecipitate the desired protein, fusion protein, or polypeptide from solution and react with the protein, fusion protein, or polypeptide on Western blots of polyacrylamide gels.

Techniques for purifying antibodies are those which are available in the art. In a preferred embodiment, antibodies are affinity purified by passing the antibodies over a column to which amino acid sequences of the invention are bound. The bound antibody is then eluted, for example using a buffer with a high salt concentration. Any such technique may be chosen to purify antibodies of the invention.

The invention also provides DNA constructs, for expressing all or a portion of a protein of the invention in a host cell. The DNA construct comprises a promoter which is functional in the particular host cell selected. The skilled artisan can readily select an appropriate promoter from the large number of cell type-specific promoters known and used in the art. The DNA construct can also contain a transcription terminator which is functional in the host cell.

The expression construct comprises a polynucleotide segment which encodes all or a portion of a human protein encoded by SEQ ID NOs:1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, and 19 or a variant thereof. The polynucleotide segment is located downstream from the promoter. Transcription of the polynucleotide segment initiates at the promoter. DNA constructs can be linear or circular and can contain sequences, if desired, for autonomous replication.

The host cell comprising the DNA construct can be any suitable prokaryotic or eukaryotic cell. Expression systems in bacteria include those described in Chang

et al., *Nature* (1978) 275: 615; Goeddel *et al.*, *Nature* (1979) 281: 544; Goeddel *et al.*, *Nucleic Acids Res.* (1980) 8: 4057; EP 36,776; U.S. 4,551,433; deBoer *et al.*, *Proc. Natl. Acad. Sci. USA* (1983) 80: 21-25; and Siebenlist *et al.*, *Cell* (1980) 20: 269.

5 Expression systems in yeast include those described in Hinnen *et al.*, *Proc. Natl. Acad. Sci. USA* (1978) 75: 1929; Ito *et al.*, *J. Bacteriol.* (1983) 153: 163; Kurtz *et al.*, *Mol. Cell. Biol.* (1986) 6: 142; Kunze *et al.*, *J. Basic Microbiol.* (1985) 25: 141; Gleeson *et al.*, *J. Gen. Microbiol.* (1986) 132: 3459, Roggenkamp *et al.*, *Mol. Gen. Genet.* (1986) 202 :302); Das *et al.*, *J. Bacteriol.* (1984) 158: 1165; De Louvencourt *et al.*, *J. Bacteriol.* (1983) 154: 737, Van den Berg *et al.*, *Bio/Technology* (1990) 8: 135; Kunze *et al.*, *J. Basic Microbiol.* (1985) 25: 141; Cregg *et al.*, *Mol. Cell. Biol.* (1985) 5: 3376; U.S. 4,837,148; U.S. 4,929,555; Beach and Nurse, *Nature* (1981) 300: 706; Davidow *et al.*, *Curr. Genet.* (1985) 10: 380; Gaillardin *et al.*, *Curr. Genet.* (1985) 10: 49; Ballance *et al.*, *Biochem. Biophys. Res. Commun.* (1983) 112: 284-289; Tilburn *et al.*, *Gene* (1983) 26: 205-22; Yelton *et al.*, *Proc. Natl. Acad. Sci. USA* (1984) 81: 1470-1474; Kelly and Hynes, *EMBO J.* (1985) 4: 475479; EP 244,234; and WO 91/00357.

15 Expression of heterologous genes in insects can be accomplished as described in U.S. 4,745,051; Friesen *et al.* (1986) "The Regulation of Baculovirus Gene Expression" in: THE MOLECULAR BIOLOGY OF BACULOVIRUSES (W. Doerfler, ed.); EP 127,839; EP 155,476; Vlak *et al.*, *J. Gen. Virol.* (1988) 69: 765-776; Miller *et al.*, *Ann. Rev. Microbiol.* (1988) 42: 177; Carbonell *et al.*, *Gene* (1988) 73: 409; Maeda *et al.*, *Nature* (1985) 315: 592-594; Lebacq-Verheyden *et al.*, *Mol. Cell. Biol.* (1988) 8: 3129; Smith *et al.*, *Proc. Natl. Acad. Sci. USA* (1985) 82: 8404; Miyajima *et al.*, *Gene* (1987) 58: 273; and Martin *et al.*, *DNA* (1988) 7:99. Numerous baculoviral strains and variants and corresponding permissive insect host cells from hosts are described in Luckow *et al.*, *Bio/Technology* (1988) 6: 47-55, Miller *et al.*, in GENERIC ENGINEERING (Setlow, J.K. *et al.* eds.), Vol. 8 (Plenum Publishing, 1986), pp. 277-279; and Maeda *et al.*, *Nature*, (1985) 315: 592-594.

25 Mammalian expression can be accomplished as described in Dijkema *et al.*,

EMBO J. (1985) 4: 761; Gorman *et al.*, *Proc. Natl. Acad. Sci. USA* (1982b) 79: 6777; Boshart *et al.*, *Cell* (1985) 41: 521; and U.S. 4,399,216. Other features of mammalian expression can be facilitated as described in Ham and Wallace, *Meth. Enz.* (1979) 58: 44; Barnes and Sato, *Anal. Biochem.* (1980) 102: 255; U.S. 4,767,704; U.S. 4,657,866; U.S. 4,927,762; U.S. 4,560,655; WO 90/103430, WO 87/00195, and U.S. RE 30,985.

DNA constructs of the invention can be introduced into host cells using any technique known in the art. These techniques include transferrin-polycation-mediated DNA transfer, transfection with naked or encapsulated nucleic acids, liposome-mediated cellular fusion, intracellular transportation of DNA-coated latex beads, protoplast fusion, viral infection, electroporation, and calcium phosphate-mediated transfection.

Alternatively, expression of an endogenous gene encoding a protein of the invention can be manipulated by introducing by homologous recombination a DNA construct comprising a transcription unit in frame with the endogenous gene, to form a homologously recombinant cell comprising the transcription unit. The transcription unit comprises a targeting sequence, a regulatory sequence, an exon, and an unpaired splice donor site. The new transcription unit can be used to turn the endogenous gene on or off as desired. This method of affecting endogenous gene expression is taught in U.S. 5,641,670, which is incorporated herein by reference.

The targeting sequence is a segment of at least 10, 12, 15, 20, or 50 contiguous nucleotides selected from the nucleotide sequences shown in SEQ ID NOs: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, and 19. The transcription unit is located upstream to a coding sequence of the endogenous gene. The exogenous regulatory sequence directs transcription of the coding sequence of the endogenous gene.

Secreted proteins of the invention have a variety of uses. For example, secreted proteins can be used in assays to determine biological activities, such as cytokine, cell proliferation, or cellular differentiation activities, tissue growth or

regeneration, activin or inhibin activity, chemotactic or chemokinetic activity, hemostatic or thrombolytic activity, receptor/ligand activity, tumor inhibition, or anti-inflammatory activity. Assays for these activities are known in the art and are disclosed, for example, in U.S. 5,654,173, which is incorporated herein by reference.

Proteins of the invention can also be used as biomarkers, to identify tissues or cell types which express the proteins, or a stage- or disease-specific alteration in protein expression. Proteins of the invention can be used in protein interaction assays, to identify ligands or binding proteins. Compounds which affect the biological activities of the secreted proteins or their ability to interact with specific ligands can be identified using proteins of the invention in screening assays. Proteins and antibodies of the invention can also be used to design diagnostic tests and therapeutic compositions for diseases which may be associated with altered expression of these proteins. Fusion proteins comprising, for example, signal sequences or transmembrane domains of the disclosed proteins, can be used to target other protein domains to cellular locations in which the domains are not normally found, such as bound to a cellular membrane or secreted extracellularly.

Further objects, features, and advantages of the present invention will readily occur to the skilled artisan provided with the disclosure above.

SYNOPSIS OF THE INVENTION

1. An isolated and purified human protein having an amino acid sequence selected from the group consisting of the amino acid sequences shown in SEQ ID Nos:20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, and 38.

2. An isolated and purified human protein having an amino acid sequence which is at least 85% identical to an amino acid sequence selected from the group consisting of the amino acid sequences shown in SEQ ID Nos:20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, and 38.

3. The isolated and purified human protein of item 2 wherein the amino acid sequence is at least 90% identical.

4. The isolated and purified human protein of item 2 wherein the amino acid sequence is at least 95% identical.

5. The isolated and purified human protein of item 2 wherein the amino acid sequence is at least 98% identical.

6. An isolated and purified human polypeptide comprising at least 6 contiguous amino acids of an amino acid sequence selected from the group consisting of the amino acid sequences shown in SEQ ID Nos:20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, and 38.

7. A fusion protein comprising a first protein segment and a second protein segment fused together by means of a peptide bond, wherein the first protein segment consists of at least 6 contiguous amino acids selected from the group consisting of the amino acid sequences shown in SEQ ID Nos:20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, and 38.

8. A preparation of antibodies which specifically bind to the human protein of item 1.

9. The preparation of antibodies of item 8 wherein the antibodies are monoclonal.

10. The preparation of antibodies of item 8 wherein the antibodies are polyclonal.

11. The preparation of antibodies of item 8 wherein the antibodies are single chain antibodies.

12. An isolated and purified subgenomic polynucleotide having a nucleotide sequence selected from the group consisting of the nucleotide sequences shown in SEQ ID NOs:1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, and 19.

13. An isolated and purified subgenomic polynucleotide consisting of at least 10 contiguous nucleotides of a nucleotide sequence selected from the group

consisting of the nucleotide sequences shown in SEQ ID NOs:1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, and 19.

14. An isolated gene corresponding to a cDNA sequence selected from the group consisting of the nucleotide sequences shown in SEQ ID NOs:1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, and 19.

15. A DNA construct for expressing all or a portion of a human protein having an amino acid sequence selected from the group consisting of the amino acid sequences shown in SEQ ID Nos:20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, and 38, comprising:

a promoter; and

a polynucleotide segment encoding at least 6 contiguous amino acids of the human protein, wherein the polynucleotide segment is located downstream from the promoter, wherein transcription of the polynucleotide segment initiates at or 3' to the promoter.

16. A host cell comprising a DNA construct comprising:

a promoter; and

a polynucleotide segment encoding at least 6 contiguous amino acids of a human protein having an amino acid sequence selected from the group consisting of the amino acid sequences shown in SEQ ID Nos:20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, and 38, wherein the polynucleotide segment is located downstream from the promoter and wherein transcription of the polynucleotide segment initiates at or 3' to the promoter.

17. A homologously recombinant cell having incorporated therein a new transcription initiation unit, wherein the new transcription initiation unit comprises in 5' to 3' order:

(a) an exogenous regulatory sequence;

(b) an exogenous exon; and

(c) a splice donor site,

wherein the transcription initiation unit is located upstream to a coding sequence of a gene, wherein the gene comprises a nucleotide sequence selected from the group

consisting of the nucleotide sequences shown in SEQ ID NOs:1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, and 19, and wherein the exogenous regulatory sequence controls transcription of the coding sequence of the gene.

18. A method of producing a human protein, comprising the steps of:

5 growing a culture of a cell comprising a DNA construct comprising (1) a promoter and (2) a polynucleotide segment encoding at least 6 contiguous amino acids of a human protein having an amino acid sequence selected from the group consisting of the amino acid sequences shown in SEQ ID Nos:20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, and 38, wherein the
10 polynucleotide segment is located downstream from the promoter and wherein transcription of the polynucleotide segment initiates at or 3' to the promoter; and; purifying the protein from the culture.

19. A method of producing a human protein, comprising the steps of:

15 growing a culture of a homologously recombinant cell having incorporated therein a new transcription initiation unit, wherein the new transcription initiation unit comprises in 5' to 3' order:

- (a) an exogenous regulatory sequence;
- (b) an exogenous exon; and
- (c) a splice donor site,

20 wherein the transcription initiation unit is located upstream to a coding sequence of a gene, wherein the gene comprises a nucleotide sequence selected from the group consisting of the nucleotide sequences shown in SEQ ID NOs:1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, and 19 and wherein the exogenous regulatory sequence controls transcription of the coding sequence of the gene; and

25 purifying the protein from the culture.

20. A method of identifying a secreted polypeptide which is modified by rough microsomes, comprising the steps of:

transcribing *in vitro* a population of cDNA molecules whereby a population of cRNA molecules is formed;

translating a first portion of the population of cRNA molecules *in vitro* in the absence of rough microsomes whereby a first population of polypeptides is formed;

5 translating a second portion of the population of cRNA molecules *in vitro* in the presence of rough microsomes whereby a second population of polypeptides is formed;

comparing the first population of polypeptides with the second population of polypeptides; and

10 detecting polypeptide members of the second population which have been modified by the rough microsomes.

21. The method of item 20 wherein the population of cDNA molecules is synthesized by reverse transcription of a population of mRNA molecules.

22. The method of item 21 wherein the mRNA molecules are isolated from a mammal.

15 23. The method of item 22 wherein the mRNA molecules are isolated from a human.

24. The method of item 20 wherein the population of cDNA molecules is obtained from a cDNA library.

20 25. The method of item 24 wherein the cDNA library is derived from a mammalian genome.

26. The method of item 25 wherein the cDNA library is derived from a human genome.

SEQUENCE LISTING

(1) GENERAL INFORMATION

(i) APPLICANT: Chiron Corporation

(ii) TITLE OF THE INVENTION: Secreted Human Proteins

(iii) NUMBER OF SEQUENCES: 38

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: Banner & Witcoff

(B) STREET: 1001 G Street, NW

(C) CITY: Washington

(D) STATE: DC

(E) COUNTRY: USA

(F) ZIP: 20001

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Diskette

(B) COMPUTER: IBM Compatible

(C) OPERATING SYSTEM: DOS

(D) SOFTWARE: FastSEQ for Windows Version 2.0

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER:

(B) FILING DATE: 11-DEC-1997

(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: 60/032757

(B) FILING DATE: 11-DEC-1996

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: Kagan, Sarah A

(B) REGISTRATION NUMBER: 32141

(C) REFERENCE/DOCKET NUMBER:

2441.39505;1369.002;1452.001

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: 202-508-9100

(B) TELEFAX: 202-508-9299

(C) TELEX:

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2063 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ix) FEATURE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

GAATTCGGCA CGAGGCCTCA GTCTTCCAGG GCGGCGGTGG GTGTCCGCTT CTCTCTGCTC	60
TTCGACTGCA CCGCACTCGC GCGTGACCCT GACTCCCCCT AGTCAGCTCA GCGGTGCTGC	120
CATGGCGTGG CCGCGGCGCG AAGCCGGCGT CCGGGCTCGC GGCGTGTTGG CTCTGGCGTT	180
GCTCGCCCTG GCCCTGTGCG TGCCCGGGGC CCGGGGCCGG GCTCTCGAGT GGTTCCTCGC	240

CGTGGTAAAC ATCGAGTACG TGGACCCGCA GACCAACCTG ACGGTGTGGA GCGTCTCGGA	300
GAGTGGCCGC TTCGGCGACA GCTCGCCCAA GGAGGGCGCG CATGGCCTGG TGGGCGTCCC	360
GTGGGCGCCC GCGGAGACC TCGAGGGCTG CGCGCCCGAC ACGCGCTTCT TCGTGCCCGA	420
GCCCCGCGGC CGAGGGGCGG CGCCCTGGGT CGCCCTGGTG GCTCGTGGGG GCTGCACCTT	480
CAAGGACAAG GTGCTGGTGG CGGCGCGGAG GAACGCCTCG GCCGTCGTCC TCTACAATGA	540
GGAGCGCTAC GGAACATCA CCTTGCCCAT GTCTCACGCG GGAACAGGAA ATATAGTGGT	600
CATTATGATT AGCTATCCAA AAGGAAGAGA AATTTTGGAG CTGGTGCAA AAGGAATTCC	660
AGTAACGATG ACCATAGGGG TTGGCACCCG GCATGTACAG GAGTTCATCA GCGGTCAGTC	720
TGTGGTGTTC GTGGCCATTG CCTTCATCAC CATGATGATT ATCTCGTTAG CCTGGCTAAT	780
ATTTTACTAT ATACAGCGTT TCCTATATAC TGGCTCTCAG ATTGGAAGTC AGAGCCATAG	840
AAAAGAACT AAGAAAGTTA TTGGCCAGCT TCTACTTCAT ACTGTAAAGC ATGGAGAAAA	900
GGGAATTGAT GTTGATGCTG AAAATTGTGC AGTGTGTATT GAAAATTTCA AAGTAAAGGA	960
TATTATTAGA ATTCTGCCAT GCAAGCATAT TTTTCATAGA ATATGCATTG ACCCATGGCT	1020
TTTGATCAC CGAACATGTC CAATGTGTAA ACTTGATGTC ATCAAAGCCC TAGGATATTG	1080
GGGAGAGCCT GGGGATGTAC AGGAGATGCC TGCTCCAGAA TCTCCTCCTG GAAGGGATCC	1140
AGCTGCAAAT TTGAGTCTAG CTTTACCAGA TGATGACGGA AGTGATGACA GCAGTCCACC	1200
ATCAGCCTCC CCTGCTGAAT CTGAGCCACA GTGTGATCCC AGCTTTAAAG GAGATGCAGG	1260
AGAAAATACG GCATTGCTAG AAGCCGGCAG GAGTGACTCT CGGCATGGAG GACCCATCTC	1320
CTAGCACACG TGCCCACTGA AGTGGCACCA ACAGAAGTTT GGCTTGAAC AAAGGACATT	1380
TTATTTTTTT TACTTTAGCA CATAATTTGT ATATTTGAAA ATAATGTATA TTATTTTACC	1440
TATTAGATTC TGATTTGATA TACAAAGGAC TAAGATATTT TCTTCTTGAA GAGACTTTTC	1500
GATTAGTGGT CATATATTTA TCTACTAAAA TAGAGTGTTC ACCATGAACA GTGTGTTGCT	1560
TCAGACTATT ACAAAGACAA CTGGGGCAGG TACTCTAATA TAAAGGACAG GTGGTGTTC	1620
TAAATAATTG GCTGCTATGG TTCTGTAAAA ACCAGTTAAT TCTATTTTC AAGGTTTTTG	1680
GCAAAGCACA TCAATGTTAG ACTAGTTGAA GTGGAATTGT ATAATTCAT TCGATAATTG	1740
ATCTCATGGG CTTTCCCTGG AGGAAAGGTT TTTTTTGTG TTTTTTTTTT AAGAACTTGA	1800
AACTTGTAAG CTGAGATGTC TGTAGCTTTT TTGCCATCT GTAGTGTATG TGAAGATTTT	1860
AAAACCTGAG AGCACTTTT CTTTGTTTAG AATTATGAGA AAGGCACTAG ATGACTTTAG	1920
GATTTGCATT TTTCCCTTTA TTGCCTCATT TCTTGTGACG CCTTGTGGG GAGGGAAATC	1980
TGTTTATTTT TTCCTACAAA TAAAAAGCTA AGATTCTATA TCGCAAAAAA AAAAAAAAAA	2040
AAAAAAAAA TTCCTGCGGC CGC	2063

(2) INFORMATION FOR SEQ ID NO:2:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1328 base pairs

- (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

GAATTCGGCA CGAGGTAGGC AAGGGATAAA AAGGCACCTA AGGCCCTTTT GCAATAAGAA	60
GCCAGATGGA TAAAGGAAGT GCTGGTCACC CTGGAGGTGT ACTGGTTTGG GGAAGGTCCC	120
CGGCCCCCAC AGCCCTCTGG GGAGCCTCAC CCTGGCTCTC CCCACTCACC TCAGCCCTCA	180
GGCAGCCCTT CCACAGGGCC CCTCTCCTGC CTGGACAGCT CTGCTGGTCT CCCCCTCCCC	240
TGGAGAAGAA CAAGGCCATG GGTGCGCCCC TGCTGCTGCC CCTGCTGCTC CTGCTGCAGC	300
CGCCAGCATT TCTGCAGCCT GGTGGCTCCA CAGGATCTGG TCCAAGCTAC CTTTATGGGG	360
TCACTCAACC AAAACACCTC TCAGCCTCCA TGGGTGGCTC TGTGGAAATC CCCTTCTCCT	420
TCTATTACCC CTGGGAGTTA GCCATAGTTC CCAACGTGAG AATATCCTGG AGACGGGGCC	480
ACTTCCACGG GCAGTCCTTC TACAGCACAA GGCCGCCTTC CATTACAAG GATTATGTGA	540
ACCGGCTCTT TCTGAACTGG ACAGAGGGTC AGGAGAGCGG CTTCTCTCAGG ATCTCAAACC	600
TGCGGAAGGA GGACCAGTCT GTGTATTTCT GCCGAGTCGA GCTGGACACC CGGAGATCAG	660
GGAGGCAGCA GTTGCACTCC ATCAAGGGGA CCAAACCTCAC CATCACCCAG GCTGTCACAA	720
CCACCACCAC CTGGAGGCCC AGCAGCACAA CCACCATAGC CGGCCTCAGG GTCACAGAAA	780
GCAAAGGGCA CTCAGAATCA TGGCACCTAA GTCTGGACAC TGCCATCAGG GTTGCAATGG	840
CTGTGCTGT GCTCAAAACT GTCATTTTGG GACTGCTGTG CCTCCTCCTC CTGTGGTGGA	900
GGAGAAGGAA AGGTAGCAGG GCGCCAAGCA GTGACTTCTG ACCAACAGAG TGTGGGGAGA	960
AGGGATGTGT ATTAGCCCCG GAGGACGTGA TGTGAGACCC GCTTGTGAGT CCTCCACACT	1020
CGTTCCCCAT TGGCAAGATA CATGGAGAGC ACCCTGAGGA CCTTTAAAAG GCAAAGCCGC	1080
AAGGCAGAAG GAGGCTGGGT CCCTGAATCA CCGACTGGAG GAGAGTTACC TACAAGAGCC	1140
TTCATCCAGG AGCATCCACA CTGCAATGAT ATAGGAATGA GGTCTGAACT CCACTGAATT	1200
AAACCACTGG CATTTGGGGG CTGTTTATTA TAGCAGTGCA AAGAGTTCCT TTATCTCCC	1260
CAAGGATGGA AAAATACAAT TTATTTTGCT TACCATAAAA AAAAAAAAAA AAAAATTCCT	1320
GCGGCCGC	1328

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1689 base pairs
 (B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

GAATTTCGGCA	CGAGGGCAAG	ATTCGATACA	AAACCAATGA	ACCTGTGTGG	GAGGAAAAC	60
TCACTTTCTT	CATTACAAAT	CCCAAGCGCC	AGGACCTTGA	AGTTGAGGTC	AGAGACGAGC	120
AGCACCAGTG	TTCCCTGGGG	AACCTGAAGG	TCCCCCTCAG	CCAGCTGCTC	ACCAGTGAGG	180
ACATGACTGT	GAGCCAGCGC	TTCCAGCTCA	GTAACCTCGG	TCCAAACAGC	ACCATCAAGA	240
TGAAGATTGC	CCTGCGGGTG	CTCCATCTCG	AAAAGCGAGA	AAGGCCTCCA	GACCACCAAC	300
ACTCAGCTCA	AGTCAAACGT	CCCTCTGTGT	CCAAAGAGGG	GAGGAAAACA	TCCATCAAAT	360
CTCATATGTC	TGGGTCTCCA	GGCCCTGGTG	GCAGCAACAC	AGCTCCATCC	ACACCAGTCA	420
TTGGGGGCAG	TGATAAGCCT	GGTATGGAAG	AAAAGGCCCA	GCCCCCTGAG	GCCGGCCCTC	480
AGGGGCTGCA	CGACCTGGGC	AGAAGCTCCT	CCAGCCTCCT	GGCCTCCCCA	GGCCACATCT	540
CAGTCAAGGA	GCCGACCCCC	AGCATCGCCT	CGGACATCTC	GCTGCCCATC	GCCACCCAGG	600
AGCTGCGGCA	AAGGCTGAGG	CAGCTGGAAA	ACGGGACGAC	CCTGGGACAG	TCTCCACTGG	660
GGCAGATCCA	GCTGACCATC	CGGCACAGCT	CGCAGAGAAA	CAAGCTTATC	GTGGTCGTGC	720
ATGCCTGCAG	AAACCTCATT	GCCTTCTCTG	AAGACGGCTC	TGACCCCTAT	GTCCGCATGT	780
ATTTATTACC	AGACAAGAGG	CGGTCAGGAA	GGAGGAAAAC	ACACGTGTCA	AAGAAAACAT	840
TAAATCCAGT	GTTTGATCAA	AGCTTTGATT	TCAGTGTTTC	GTTACCAGAA	GTGCAGAGGA	900
GAAEGCTEGA	CGTTGEEGTG	AAGAACAGTG	GCGGCTTECT	GTECAAGAC	AAAGGGETCC	960
TTGGCAAAGT	ATTGGTTGCT	CTGGCATCTG	AAGAACTTGC	CAAAGGCTGG	ACCCAGTGGT	1020
ATGACCTCAC	GGAAGATGGG	ACGAGGCCTC	AGGCGATGAC	ATAGCCGAG	CAGGCAGGAG	1080
GCGTCCTCTT	CAGCGTAGCT	CTCCACCTCT	ACCCGGAACA	CACCTCTCA	CAGACGTACC	1140
AATGTTATTT	TTATAATTTT	ATGGATTAG	TTATACATAC	CTTAATAGTT	TTATAAAATT	1200
GTTGACATTT	CAGGCAAATT	TGGCCAATAT	TATCATTGAA	TTTTCTGTGT	TGGATTTCTT	1260
CTAGGATTTT	GCCAGTTCCT	ACAACGTGCA	GTAGGGCGGC	GGTAGCTCTT	GTGTCTGTGG	1320
ACTCTGCTCA	GCTGTGTCCG	TAGGAGTCGG	ATGTGTCTGT	GCTTTATTAT	GGCCTTGTTT	1380
ATATATCACT	GAGGTATACT	ATGCCATGTA	AATAGACTAT	TTTTTATAAT	CTTAACATGC	1440
TGGTTTAAAT	TCAGAAGGAA	ATAGATCAAG	GAAATATATA	TATTTTCTTC	TAAAACTTAT	1500
TAAATTCGTG	TGACAAATAA	TCATTTTCAT	CTTGGCAGCA	AAAAGTTCTC	AGTGACCTAT	1560
TTTGTGGTGT	TTCTTTTGA	AAAGAAAAGC	TGAAATATTA	TTAAATGCTA	GTATGTTTCT	1620
GCCCATTATG	AAAGATGAAA	TAAAGTATTC	AAAATATTAA	AAAAAAAAAA	AAAAAATTCC	1680
TGCGGCCGC						1689

(2) INFORMATION FOR SEQ ID NO:4:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1505 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

GAATTCGGCA	CGAGGAGCAG	ATCTGCAAGA	GTTTCGTTTA	TGGAGGCTGC	TTGGGCAACA	60
AGAACAATA	CCTTCGGGAA	GAAGAGTGCA	TTCTAGCCTG	TCGGGGTGTG	CAAGGTGGGC	120
CTTTGAGAGG	CAGCTCTGGG	GCTCAGGCGA	CTTTCCCCCA	GGGCCCCTCC	ATGGAAAGGC	180
GCCATCCAGT	GTGCTCTGGC	ACCTGTCAGC	CCACCCAGTT	CCGCTGCAGC	AATGGCTGCT	240
GCATCGACAG	TTTCCTGGAG	TGTGACGACA	CCCCCAACTG	CCCCGACGCC	TCCGACGAGG	300
CTGCCTGTGA	AAAATACACG	AGTGGCTTTG	ACGAGCTCCA	GCGCATCCAT	TTCCCCAGCG	360
ACAAAGGGCA	CTGCGTGGAC	CTGCCAGACA	CAGGACTCTG	CAAGGAGAGC	ATCCCGCGCT	420
GGTACTACAA	CCCCTTCAGC	GAACACTGCG	CCCCTTTTAC	CTATGGTGGT	TGTTACGGCA	480
ACAAGAACAA	CTTTGAGGAA	GAGCAGCAGT	GCCTCGAGTC	TTGTCGCGGC	ATCTCCAAGA	540
AGGATGTGTT	TGGCCTGAGG	CGGGAAATCC	CCATTCCCAG	CACAGGCTCT	GTGGAGATGG	600
CTGTGCGCAGT	GTTCTCTGGT	ATCTGCATTG	TGGTGGTGGT	AGCCATCTTG	GGTTACTGCT	660
TCTTCAAGAA	CCAGAGAAAG	GACTTCCACG	GACACCACCA	CCACCCACCA	CCCACCCCTG	720
CCAGCTCCAC	TGTCTCCACT	ACCGAGGACA	CGGAGCACCT	GGTCTATAAC	CACACCACGC	780
GGCCCCCTCTG	AGCCTGGGTC	TCACCGGCTC	TCACCTGGCC	CTGCTTCCTG	CTTGCCAAGG	840
CAGAGGCCTG	GGCTGGGAAA	AACTTTGGAA	CCAGACTCTT	GCCTGTTTCC	CAGGCCCACT	900
GTGCCTCAGA	GACCAGGGCT	CCAGCCCCCTC	TTGGAGAAGT	CTCAGCTAAG	CTCACGTCCT	960
GAGAAAGCTC	AAAGGTTTGG	AAGGAGCAGA	AAACCCTTGG	GCCAGAAGTA	CCAGACTAGA	1020
TGGACCTGCC	TGCATAGGAG	TTTGGAGGAA	GTTGGAGTTT	TGTTTCCTCT	GTTCAAAGCT	1080
GCCTGTCCCT	ACCCCATGGT	GCTAGGAAGA	GGAGTGGGGT	GGTGTGAGAC	CCTGGAGGCC	1140
CCAACCCTGT	CCTCCCGAGC	TCCTCTTCCA	TGCTGTGCGC	CCAGGGCTGG	GAGGAAGGAC	1200
TTCCCTGTGT	AGTTTGTGCT	GTAAAGAGTT	GCTTTTGTGT	TATTTAATGC	TGTGGCATGG	1260
GTGAAGAGGA	GGGGAAGAGG	CCTGTTTGGC	CTCTCTATCC	TCTCTTCCTC	TTCCCCAAG	1320
ATTGAGCTCT	CTGCCCTTGA	TCAGCCCCAC	CCTGGCCTAG	ACCAGCAGAC	AGAGCCAGGA	1380
GAAGCTCAGC	TGCATTCCGC	AGCCCCCACC	CCCAAGGTTT	TCCAACATCA	CAGCCCAGCC	1440
CGCCCACTGG	GTAATAAAAG	TGTTTGTGG	AAAAAAAAAA	AAAAAAAAAA	AAGTCCTGCG	1500

GCCGC

1505

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2002 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

GAATTCGGCA CGAGGGCCAT GCCCGGGCTA TCCCGCGGGT CCGCGCGCGC ACTGCTCGCC	60
GCCCTGCTGG CGTCGACGCT GTTGGCGCTG CTCGTGTGCG CCGCGCGGGG TCGCGGCGGC	120
CGGGACCACG GGGACTGGGA CGAGGCCTCC CGGCTGCCGC CGCTACCACC CCGCGAGGAC	180
GCGGCGCGCG TGGCCCGCTT CGTGACGCAC GTCTCCGACT GGGGCGCTCT GGCCACCATC	240
TCCACGCTGG AGGCGGTGCG CGGCCGGCCC TTCGCCGACG TCCTCTCGCT CAGCGACGGG	300
CCCCCGGGCG CGGGCAGCGG CGTGCCCTAT TTCTACCTGA GCGCGCTGCA GCTCTCCGTG	360
AGCAACCTGC AGGAGAATCC ATATGCTACA CTGACCATGA CTTTGGCACA GACCAACTTC	420
TGCAAGAAAC ATGGATTGTA TCCACAAAGT CCCCTTTGTG TTCACATAAT GCTGTCAGGA	480
ACTGTGACCA AGGTGAATGA AACAGAAATG GATATTGCAA AGCAATTCGT ATTCAATCGA	540
CACCCTGAGA TGAAAACCTG GCCTTCCAGC CATAATTGGT TCTTTGCTAA GTTGAATATA	600
ACCAATATCT GGGTCCTGGA CTACTTTGGT GGACCAAAAA TCGTGACACC AGAAGAATAT	660
TATAATGTCA CAGTTCAGTG AAGCAGACTG TGGTGAATTT AGCAACACTT ATGAAGTTTC	720
TTAAAGTGGC TCATACACAC TTAAAAGGCT TAATGTTTCT CTGGAAGCG TCCCAGAATA	780
TTAGCCAGTT TTCTGTCACA TGCTGGTTTG TTTGCTTGCT TGTTTACTTG CTTGTTTACC	840
AATAGAGTTG ACCTGTTATT GGATTTCCTG GAAGATGTGG TAGCTACTTT TTTCTATTTT	900
TGAAGCCATT TTCGTAGAGA AATATCCTTC ACTATAATCA AATAAGTTTT GTCCCATCAA	960
TTCCAAAGAT GTTTCAGTG GTGCTCTTGA AGAGGAATGA GTACCAGTTT TAAATTGCCC	1020
ATTGGCATTG GAAGGTAGTT GAGTATGTGT TCTTTATTCC TAGAAGCCAC TGTGCTTGGT	1080
AGAGTGCATC ACTCACCACA GCTGCCTCTT GAGCTGCCTG AGCCTGGTGC AAAAGGATTG	1140
GCCCCATTA TGGTGCTTCT GAATAAATCT TGCCAAGATA GACAAACAAT GATGAACTC	1200
AGATGGAGCT TCCTACTCAT GTTGATTTAT GTCTCACAAT CCTGGGTATT GTTAATTCAA	1260
CATAGGGTGA AACTATTTCT GATAAAGAAC TTTTGAAAAA CTTTTTATAC TCTAAAGTGA	1320
TACTCAGAAC AAAAGAAAGT CATAAACTC CTGAATTTAA TTTCCCACC TAAGTCGAGA	1380

CAGTATTATC	AAAACACATG	TGCACACAGA	TTATTTTTTG	GCTCCAAAAC	TGGATTGCAA	1440
AAGAAAGAGG	AGAGATATTT	TGTGTGTTCC	TGGTATTCTT	TTATAAGTAA	AGTTACCCAG	1500
GCATGGACCA	GCTTCAGCCA	GGGACAAAAT	CCCCTCCCAA	ACCACTCTCC	ACAGCTTTTT	1560
AAAAATACTT	CTACTCTTAA	CAATTACCTA	AGGTTCCCTC	AAACCCCCC	AACTCTTAAT	1620
AGCTTCTAGT	GCTGCTACAA	TCTAAGTCAG	GTCACCAGAG	GGAAGAGAAC	ATGGCATTAA	1680
AAGAATCACA	TCTTCAGAAG	AGAAGACACT	AATATTATTA	CCCATATACA	TGATTTCAGA	1740
AGATGACATA	AGATTCCTCT	TAAAGAGGAA	ATGTCAGGAA	TCAAGCCACT	GAATCCTTAA	1800
AGAGAAAAGT	TGAATATGAG	TCATTGTGTC	TGAAACTGTC	AAAGTGAAGT	TAAGTGAGAT	1860
CCAGCAAACA	GGTTCTGTTT	AAGAAAATA	ATTTATACTA	AATTTAGTAA	AATGGACTTC	1920
TTATTCAAAAG	CATCAATAAT	TAAAAGAATT	ATTTTAAAAA	AAAAAAAAAA	AAAAAAAAAA	1980
AAAAAAAAAAT	TCCTGCGGCC	GC				2002

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1322 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

GAATTCGGCA	CGAGGGCCAC	GACTCTGCTG	GCATTTCTTC	TATAGCCACT	GGAATCTGAT	60
CCTGATTGTC	TTCCACTACT	ACCAGGCCAT	CACCACTCCG	CCTGGGTACC	CACCCAGGG	120
CAGGAATGAT	ATCGCCACCG	TCTCCATCTG	TAAGAAGTGC	ATTTACCCCA	AGCCAGCCCG	180
AACACACCAC	TGCAGCATCT	GCAACAGGTG	TGTGCTGAAG	ATGGATCACC	ACTGCCCCTG	240
GCTAAACAAT	TGTGTGGGCC	ACTATAACCA	TCGGTACTTC	TTCTCTTTCT	GCTTTTTTCAT	300
GACTCTGGGC	TGTGTCTACT	GCAGCTATGG	AAGTTGGGAC	CTTTTCCGGG	AGGCTTATGC	360
TGCCATTGAG	AAAATGAAAC	AGCTCGACAA	GAACAACTA	CAGGCGGTTG	CCAACCAGAC	420
TTATCACCAG	ACCCACCAC	CCACCTTCTC	CTTTCGAGAA	AGGATGACTC	ACAAGAGTCT	480
TGTCTACCTC	TGGTTCCTGT	GCAGTTCTGT	GGCACTTGCC	CTGGGTGCCC	TAAGTGATG	540
GCATGCTGTT	CTCATCAGTC	GAGGTGAGAC	TAGCATCGAA	AGGCACATCA	ACAAGAAGGA	600
GAGACGTCCG	CTACAGGCCA	AGGGCAGAGT	ATTTAGGAAT	CCTTACAACT	ACGGCTGCTT	660
GGACAACTGG	AAGGTATTCC	TGGGTGTGGA	TACAGGAAGG	CACTGGCTTA	CTCGGGTGCT	720
CTTACCTTCT	ACTCACTTGC	CCCATGGGAA	TGGAATGAGC	TGGGAGCCCC	CTCCCTGGGT	780

GACTGCTCAC	TCAGCCTCTG	TGATGGCAGT	GTGAGCTGGA	CTGTGTCAGC	CACGACTCGA	840
GCACTCATTG	TGCTCCCTAT	GTTATTTCAA	GGGCCTCCAA	GGGCAGCTTT	TCTCAGAATC	900
CTTGATCAAA	AAGAGCCAGT	GGGCCTGCCT	TAGGGTACCA	TGCAGGACAA	TTCAAGGACC	960
AGCCTTTTTA	CCACTGCAGA	AGAAAGACAC	AATGTGGAGA	AATCTTAGGA	CTGACATCCC	1020
TTTACTCAGG	CAAACAGAAAG	TTCCAACCCC	AGACTAGGGG	TCAGGCAGCT	AGCTACCTAC	1080
CTTGCCCAGT	GCTGACCCGG	ACCTCCTCCA	GGATACAGCA	CTGGAGTTGG	CCACCACCTC	1140
TTCTACTTGC	TGTCTGAAAA	AACACCTGAC	TAGTACAGCT	GAGATCTTGG	CTTCTCAACA	1200
GGGCAAAGAT	ACCAGGCCTG	CTGCTGAGGT	CACTGCCACT	TCTCACATGC	TGCTTAAGGG	1260
AGCACAAATA	AAGGTATTCT	ATTTTTAAAA	AAAAAAAAAA	AAAAAAAAAT	TCCTGCGGCC	1320
GC						1322

(2) INFORMATION FOR SEQ ID NO:7:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1573 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

GAATTCGGCA	CGAGGAGCCT	GCCTTCATCT	AGGATGGCTC	CTCTGGGCAT	GCTGCTTGGG	60
CTGCTGATGG	CCGCCTGCTT	CACCTTCTGC	CTCAGTCATC	AGAACCTGAA	GGAGTTTGCC	120
CTGACCAACC	CAGAGAAGAG	CAGCACCAAA	GAAACAGAGA	GAAAAGAAAC	CAAAGCCGAG	180
GAGGAGCTGG	ATGCCGAAGT	CCTGGAGGTG	TTCCACCCGA	CGCATGAGTG	GCAGGCCCTT	240
CAGCCAGGGC	AGGCTGTCCC	TGCAGGATCC	CACGTACGGC	TGAATCTTCA	GACTGGGGAA	300
AGAGAGGCAA	AACTCCAATA	TGAGGACAAG	TTCCGAAATA	ATTTGAAAGG	CAAAAGGCTG	360
GATATCAACA	CCAACACCTA	CACATCTCAG	GATCTCAAGA	GTGCACTGGC	AAAATTCAAG	420
GAGGGGGCAG	AGATGGAGAG	TTCAAAGGAA	GACAAGGCAA	GGCAGGCTGA	GGTAAAGCGG	480
CTCTTCCGCC	CCATTGAGGA	ACTGAAGAAA	GACTTTGATG	AGCTGAATGT	TGTCATTGAG	540
ACTGACATGC	AGATCATGGT	ACGGCTGATC	AACAAGTTCA	ATAGTTCCAG	CTCCAGTTTG	600
GAAGAGAAGA	TTGCTGCGCT	CTTTGATCTT	GAATATTATG	TCCATCAGAT	GGACAATGCG	660
CAGGACCTGC	TTTCCTTTGG	TGGTCTTCAA	GTGGTGATCA	ATGGGCTGAA	CAGCACAGAG	720
CCCCTCGTGA	AGGAGTATGC	TGCGTTTGTG	CTGGGCGCTG	CCTTTTCCAG	CAACCCCAAG	780
GTCCAGGTGG	AGGCCATCGA	AGGGGGAGCC	CTGCAGAAGC	TGCTGGTCAT	CCTGGCCACG	840

GAGCAGCCGC	TC	ACTG	CAAA	GA	GAAG	GTC	CT	GTT	GC	AC	TG	GCT	CC	CT	G	CG	CC	AC	900
TTCCCTATG	CC	CAG	CG	GCA	GT	TCT	GA	AG	CT	CG	GG	GG	GC	TG	CAG	GT	CCT	GAG	960
GTGCAGGAGA	AG	GG	CAC	GGA	GG	TG	CT	CG	CC	GT	GC	CG	TG	TC	AC	ACT	GCT	CT	1020
GTCACGGAGA	AG	AT	GT	T	CG	C	GAG	GAG	AG	GCT	GAG	CT	GA	CC	CAG	GAT	GT	CCCC	1080
AAGCTGCAGC	AG	TAT	CG	CCA	GG	TAC	AC	CT	C	CT	G	CC	AG	GC	TG	GG	AA	CA	1140
GAGATCACGG	CC	CAC	CT	CCT	GG	CG	CT	G	CCC	GAG	CAT	GAT	G	CC	CG	TG	AG	AA	1200
ACACTGGGCG	TC	CT	CT	G	AC	CT	G	CC	GG	GAC	CG	T	ACC	GT	CAG	G	AC	CC	1260
AGGACACTGG	CC	AG	CCT	G	CA	GG	CT	G	AG	TAC	CAG	GT	G	CC	AG	CCT	G	GA	1320
GGTGAGGACG	AG	GG	G	CT	ACT	T	CC	AG	G	AG	CT	G	T	CT	GA	AC	AG	CT	1380
CTGAGATGAG	GC	CC	CA	CACC	AG	GA	CT	G	G	AC	TGG	AT	G	CC	TA	GAG	GC	TG	1440
CAGCGTGGGT	GG	G	CT	T	CT	CA	GG	CAG	G	AG	GA	CAT	CT	T	G	G	CA	GT	1500
TGGAACCTG	AAG	G	CC	AAAA	AAAA	AAAA	AAAA	AAAA	AAAA	AAAA	AAAA	AAAA	AAAA	AAAA	AAAA	AAAA	AAAA	AAAA	1560
TTCCTGCGGC	CG	C																	1573

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1185 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

GAATTCGGCA	CG	AGGG	GGG	CT	TTA	AGGG	GACA	GCT	GAG	CCGG	CAG	GTGG	CAG	AT	CAG	AT	GTG	60
GCAGGCTGGG	AAA	AGACA	AG	CCT	CC	AGGG	GC	CTT	CAG	CTTG	TAC	GCCA	ACA	TC	GAC	AT	CCCT	120
CAGACCCTAC	TTT	GAT	GTGG	AG	CT	GCT	CA	GGT	GCG	AAGC	AGG	CTC	CTGG	AGT	CC	AT	GAT	180
CCCTATCAAG	AT	GGT	CAACT	T	CCCC	CAGAA	AAT	TG	CAG	GT	GAA	CT	CTATG	GAC	CT	CT	CAT	240
GCTGGTCTTC	ACT	CT	G	GTG	CT	ACT	CCAT	G	G	GAT	G	AAG	AC	G	T	CTG	ACACT	300
CCGGGAGGGC	ACC	CT	GAT	GG	GC	AC	AG	CCAT	TGG	CAC	CT	G	TTC	G	G	CT	ACT	360
CTCATCCTTC	ATT	TACT	TCC	TG	CCT	ACCT	GTG	CA	AC	GC	CAG	AT	CACCA	TG	CT	G	CAGAT	420
GTTGGCACTG	CT	GGG	CTATG	GC	CT	TTT	GG	GC	ATT	G	CATT	G	CT	G	TTC	TA	CACT	480
TATCCACCTC	CAC	G	CCCT	CT	TCT	ACCT	CTT	CT	G	G	CT	G	TTG	GT	G	G	TGGAC	540
GCGCATGGTA	GC	AG	T	GTGG	TG	TCT	CGGAC	CG	T	GGG	CCCC	AC	AC	AG	CG	GC	TGCT	600
TGGCACCCTG	GCT	G	CCCT	TAC	AC	AT	G	CT	CTT	CCT	G	CT	CTAT	CT	G	CAT	TTTG	660

AGTGGTAGAG	GGGATCCTGG	ACACACTGGA	GGGCCCCAAC	ATCCCGCCCA	TCCAGAGGGT	720
CCCCAGAGAC	ATCCCTGCCA	TGCTCCCTGC	TGCTCGGCTT	CCCACCACCG	TCCTCAACGC	780
CACAGCCAAA	GCTGTTGCGG	TGACCCTGCA	GTCACACTGA	CCCCACCTGA	AATTCTTGGC	840
CAGTCCTCTT	TCCCGCAGCT	GCAGAGAGGA	GGAAGACTAT	TAAAGGACAG	TCCTGATGAC	900
ATGTTTCGTA	GATGGGGTTT	GCAGCTGCCA	CTGAGCTGTA	GCTGCGTAAG	TACCTCCTTG	960
ATGCCTGTCT	GCACTTCTGA	AAGGCACAAG	GCCAAGAAGT	CCTGGCCAGG	ACTGCAAGGC	1020
TCTGCAGCCA	ATGCAGAAAA	TGGGTCAGCT	CCTTTGAGAA	CCCCTCCCCA	CCTACCCCTT	1080
CCTTCCTCTT	TATCTCTCCC	ACATTGTCTT	GCTAAATATA	GACTTGGTAA	TTAAAATGTT	1140
GATTGAAGTC	TGGAAAAAAA	AAAAAAAAAA	AATTCCTGCG	GCCGC		1185

(2) INFORMATION FOR SEQ ID NO:9:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1226 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

GAATTEGGGA	CGAGGGAAGC	CACCATCTTC	CTTCGGGCTG	CACCCCTTTA	AAGGCACCCA	60
GACCCCTCTG	GAAAAAGATG	AACTGAAGCC	CTTTGACATC	CTCCAGCCTA	AGGAGTACTT	120
CCAGCTCAGC	CGCCACACGG	TCATTAAGAT	GGGAAGTGAG	AACGAGGCCC	TGGATCTCTC	180
CATGAAGTCA	GTGCCCTGGC	TCAAGGCTGG	TGAAGTCAGT	CCCCCAATCT	TCCAGGAAGA	240
TGCAGCCCTA	GACCTGTCAG	TGGCAGCCCA	CCGGAATCC	GAGCCTCCCC	CTGAGACACT	300
GTATGACAGT	GGTGCATCAG	TGGACAGCTC	AGGTCACACA	GTGATGGAGA	AACTTCCCAG	360
TGGCATGGAA	ATTTCTTTTG	CCCCTGCCAC	GTCCCATGAG	GCCCCAGCCA	TGATGGATAG	420
TCACATCAGC	AGCAGTGATG	CTGCTACCGA	GATGCTCAGC	CAGCCCAACC	ACCCACGCGG	480
CGAAGTCAAG	GCTGAAAATA	ACATTGAGAT	GGTGGGCGAG	TCCCAGGCGG	CCAAGGTCAT	540
TGTCTCTGTC	GAAGATGCTG	TGCCTACCAT	ATTCTGTGGC	AAGATCAAAG	GCCTCTCAGG	600
GGTGTCCACC	AAAACTTCT	CCTTCAAAG	AGAAGACTCC	GTGCTTCAGG	GCTATGACAT	660
CAACAGCCAA	GGGGAAGAGT	CCATGGGAAA	TGCAGAGCCC	CTTAGGAAAC	CCATCAAAAA	720
CCGGAGCATA	AAGTTAAAGA	AAGTGAATC	CCAGGAAGTA	CACATGCTCC	CAATCAAAAA	780
ACAACGGCTG	GCCACCTTTT	TTCCAAGAAA	GTAAATAACG	GCTTTTAA	ATTTGTATGA	840
TTATAATATG	GGGAAAGGTG	CATTGGTTTT	ATAAAAAGGC	ATTTAAACA	AATTATCTTT	900

GTTAATTATT	TTGGGGAGTA	GTTGGGAAAT	GGAAAGGTGA	ATTGGCTCTA	GAGGCCCTGT	960
ATGCTAGTAT	CATTTTCTTT	TTTAATTTTT	GACTTTTCAC	AAATGAGTAA	ATAAGAGCAA	1020
CCTATTTTTC	AAGCAGATTG	CACATTTTTT	GCAGCTTTAA	TGGAATATTG	GGTGAATTAG	1080
AGGGGTAAAA	AAAGCTATTT	TCATTGCCAC	AAAGTGCTTT	GATGATGTAA	TACCTAATAA	1140
AGGGTAGGAT	GAATATTTCA	CAATAAATGT	TTGTTTGCAC	TAAAAAAAAA	AAAAAAAAAA	1200
AAAAAAAAAA	AAATTCCTGC	GGCCGC				1226

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1049 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

GAATTCGGCA	CGAGGGCGCC	ATGGTGAAGG	TGACGTTCAA	CTCCGCTCTG	GCCCAGAAGG	60
AGGCCAAGAA	GGACGAGCCC	AAGAGCGGCG	AGGAGGCGCT	CATCATCCCC	CCCAGCGCCG	120
TCGCGGTGGA	CTGCAAGGAC	CCAGATGATG	TGGTACCAGT	TGGCCAAAGA	AGAGCCTGGT	180
GTTGGTGCAT	GTGCTTTGGA	CTAGCATTTA	TGCTTGCAGG	TGTTATTCTA	GGAGGAGCAT	240
ACTTGTACAA	ATATTTTGCA	CTTCAACCAG	ATGACGTGTA	CTACTGTGGA	ATAAAGTACA	300
TCAAAGATGA	TGTCATCTTA	AATGAGCCCT	CTGCAGATGC	CCCAGCTGCT	CTCTACCAGA	360
CAATTGAAGA	AAATATTAAA	ATCTTTGAAG	AAGAAGAAGT	TGAATTTATC	AGTGTGCCTG	420
TCCCAGAGTT	TGCAGATAGT	GATCCTGCCA	ACATTGTTCA	TGACTTTAAC	AAGAACTTA	480
CAGCCTATTT	AGATCTTAAC	CTGGATAAGT	GCTATGTGAT	CCCTCTGAAC	ACTTCCATTG	540
TTATGCCACC	CAGAAACCTA	CTGGAGTTAC	TTATTAACAT	CAAGGCTGGA	ACCTATTTGC	600
CTCAGTCCTA	TCTGATTCAT	GAGCACATGG	TTATTACTGA	TCGCATTGAA	AACATTGATC	660
ACCTGGGTTT	CTTTATTTAT	CGACTGTGTC	ATGACAAGGA	AACTTACAAA	CTGCAACGCA	720
GAGAACTAT	TAAAGGTATT	CAGAAACGTG	AAGCCAGCAA	TTGTTTCGCA	ATTCGGCATT	780
TTGAAAACAA	ATTTGCCGTG	GAAACTTTAA	TTTGTTCTTG	AACAGTCAAG	AAAAACATTA	840
TTGAGGAAAA	TTAATATCAC	AGCATAACCC	CACCCTTTAC	ATTTTGTGTC	AGTTGATTAT	900
TTTTTAAAGT	CTTCTTTCAT	GTAAGTAGCA	AACAGGGCTT	TACTATCTTT	TCATCTCATT	960
AATTCAATTA	AAACCATTAC	CTTAAAAAAA	AAAAAAAAAA	AAAAAAAAAA	AAAAAAAAAA	1020
AAAAAAAAAA	AAAAAATTC	TGCGGCCGC				1049

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1142 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

```

GAATTCGGCA CGAGGGGAGA ATACTTTTGT CGATGCCTAC TGGAGACTTT GATTCGAAGC      60
CCAGTTGGGC CGACCAGGTG GAGGAGGAGG GGGAGGACGA CAAATGTGTC ACCAGCGAGC      120
TCCTCAAGGG GATCCCTCTG GCCACAGGTG ACACCAGCCC AGAGCCAGAG CTACTGCCGG      180
GAGCTCCACT GCCGCCTCCC AAGGAGGTCA TCAACGGAAA CATAAAGACA GTGACAGAGT      240
ACAAGATAGA TGAGGATGGC AAGAAGTTCA AGATTGTCCG CACCTTCAGG ATTGAGACCC      300
GGAAGGCTTC AAAGGCTGTC GCAAGGAGGA AGAACTGGAA GAAGTTCGGG AACTCAGAGT      360
TTGACCCCCC CGGACCCAAT GTGGCCACCA CCACTGTCAG TGACGATGTC TCTATGACGT      420
TCATCACCAG CAAAGAGGAC CTGAACTGCC AGGAGGAGGA GGACCCTATG AACAAATTCA      480
AGGGCCAGAA GATCGTGTCC TGCCGCATCT GCAAGGGCGA CCACTGGACC ACCCGCTGCC      540
CCTACAAGGA TACGCTGGGG CCCATGCAGA AGGAGCTGGC CGAGCAGCTG GGCCTGTCTA      600
CTGGCGAGAA GGAGAAGCTG CCGGGAGAGC TAGAGCCGGT GCAGGCGACG CAGAACAAGA      660
CAGGGAAGTA TGTGCCGCGG AGCCTGCGCG ACGGGGCCAG CCGCCGCGGG GAGTCCATGC      720
AGCCCAACCG CAGAGCCGAC GACAACGCCA CCATCCGTGT CACCAACTTG CGCAGAGGAC      780
ACGCGTGAGA CCGACCTGCA GGAGCTCTTC CGGCCTTTTC GCTCCATCTC CGGCATCTAC      840
CTGGCTAAGG ACAAGACCAC TGGCCAATCC AAGGGCTTTG CCTTCATCAG CTTCACCGC      900
CGCGAGGATG CTGCGCGTGC CATTGCCGGG GTGTCCGGCT TTGGCTACGA CCACCTCATC      960
CTCAACGTCG AGTGGGCCAA GCCGTCCACC AACTAAGCCA GCTGCCACTG TGTACTCGGT      1020
CCGGGACCCT TGGCGACAGA AGACAGCCTC CGAGAGCGCG GGCTCCAAGG GCAATAAAGC      1080
AGCTCCACTC TCAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAT TCCTGCGGCC      1140
GC                                                                    1142

```

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1696 base pairs

(B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

GAATTCGGCA	CGAGGGAAAC	ATGGCGGTAG	GCTGGGACCA	TAACACAAGC	ATGACTATAT	60
GAAGGAAGAG	GAAGGTTTTC	CTGAAGATGA	GGCGACTGAA	TCGGAAAAAA	ACTTTAAGTT	120
TGGTAAAAGA	GTTGGATGCC	TTCCGAAGG	TTCCTGAGAG	CTATGTAGAG	ACTTCAGCCA	180
GTGGAGGTAC	AGTTTCTCTA	ATAGCATTTA	CAACTATGGC	TTTATTAACC	ATAATGGAAT	240
TCTCAGTATA	TCAAGATACA	TGGATGAAGT	ATGAATACGA	AGTAGACAAG	GATTTTTCTA	300
GCAAAATTAAG	AATTAATATA	GATATTACTG	TTGCCATGAA	GTGTCAATAT	GTTGGAGCGG	360
ATGTATTGGA	TTTAGCAGAA	ACAATGGTTG	CATCTGCAGA	TGGTTTAGTT	TATGAACCAA	420
CAGTATTTGA	TCTTTCACCA	CAGCAGAAAG	AGTGGCAGAG	GATGCTGCAG	CTGATTCAGA	480
GTAGGCTACA	AGAAGAGCAT	TCACCTCAAG	ATGTGATATT	TAAAAGTGCT	TTTAAAAGTA	540
CATCAACAGC	TCTTCCACCA	AGAGAAGATG	ATTCATCACA	GTCTCCAAAT	GCATGCAGAA	600
TTCATGGCCA	TCTATATGTC	AATAAAGTAG	CAGGGAATTT	TCACATAACA	GTGGGCAAGG	660
CAATTCCACA	TCCTCGTGGT	CATGCACATT	TGGCAGCACT	TGTCAACCAT	GAATCTTACA	720
ATTTTCTCA	TAGAATAGAT	CATTTGTCTT	TTGGAGAGCT	TGTTCCAGCA	ATTATTAATC	780
CTTTAGATGG	AACTGAAAAA	ATTGCTATAG	ATCACAACCA	GATGTTCCAA	TATTTTATTA	840
CAGTTGTGCC	AACAAAATA	CATACATATA	AAATATCAGC	AGACACCCAT	CAGTTTTCTG	900
TGACAGAAAG	GGAACGTATC	ATTAACCATG	CTGCAGGCAG	CCATGGAGTC	TCTGGGATAT	960
TTATGAAATA	TGATCTCAGT	TCTCTTATGG	TGACAGTTAC	TGAGGAGCAC	ATGCCATTCT	1020
GGCAGTTTTT	TGTAAGACTC	TGTGGTATTG	TTGGAGGAAT	CTTTTCAACA	ACAGGCATGT	1080
TACATGGAAT	TGAAAAATTT	ATAGTTGAAA	TAATTTGCTG	TCGTTTCAGA	CTTGGATCCT	1140
ATAAACCTGT	CAATTCTGTT	CCTTTTGAGG	ATGGCCACAC	AGACAACCAC	TTACCTCTTT	1200
TAGAAAATAA	TACACATTAA	CACCTCCCGA	TTGAAGGAGA	AAAACTTTTT	GCCTGAGACA	1260
TAAACCTTT	TTTTAATAAT	AAAATATTGT	GCAATATATT	CAAAGAAAAG	AAAACACAAA	1320
TAAGCAGAAA	ACATACTTAT	TTTAAAAAAG	AAAAAAAAGG	ATAAAAAAAC	CCAAACCTGAA	1380
ATTCTATATA	CGTTGTGTCT	GTTACAAATG	TCGTAGAAGA	AATCATGCAG	CTAAACGATG	1440
AAGAAGCCCA	ACTGGAGTGT	TGCTTTGAAG	ATGACGCCTT	CTTATATTTT	CATAGCAAAT	1500
GGGTGGTATC	AAAAATCAGAC	ATTGCTTCTT	GCTGATAAAA	AGCCTGAAGG	AAATAAGTGA	1560
AACTACATCT	ATGGGAAAAA	AAAAAACATT	GAGAAGTGCA	AATGTTCGCA	TCCTTTTGTT	1620
TTTAAAAGAT	ATGATGTCAG	AATAAAATGT	GGAAAACATA	CGGAAAAAAA	AAAAAAAAAA	1680
AAATTCCTGC	GGCCGC					1696

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1100 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

GAATTCGGCA CGAGGCGGCA CGAGGCGGCA CGAGGGTGGC ATATCACGGC CATGGGGTCT	60
CAGCATTCCG CTGCTGCTCG CCCCTCCTCC TGCAGGCGAA AGCAAGAAGA TGACAGGGAC	120
GGTTTGCTGG CTGAACGAGA GCAGGAAGAA GCCATTGCTC AGTTCCCATA TGTGGAATTC	180
ACCGGGAGAG ATAGCATCAC CTGTCTCACG TGCCAGGGGA CAGGCTACAT TCCAACAGAG	240
CAAGTAAATG AGTTGGTGGC TTTGATCCCA CACAGTGATC AGAGATTGCG CCCTCAGCGA	300
ACTAAGCAAT ATGTCCTCCT GTCCATCCTG CTTTGTCTCC TGGCATCTGG TTTGGTGGTT	360
TTCTTCCTGT TTCCGCATTC AGTCCTTG TGATGATGACG GCATCAAAGT GGTGAAAGTC	420
ACATTTAATA AGCAAGACTC CCTTGTAATT CTCACCATCA TGGCCACCCT GAAAATCAGG	480
AACTCCAAC TCTACACGGT GGCAGTGACC AGCCTGTCCA GCCAGATTCA GTACATGAAC	540
ACAGTGGTCA GTACATATGT GACTACTAAC GTCTCCCTTA TTCCACCTCG GAGTGAGCAA	600
CTGGTGAATT TTACCGGGAA GGCCGAGATG GGAGGACCGT TTTCCTATGT GTACTTCTTC	660
TGCACGGTAC CTGAGATCCT GGTGCACAAC ATAGTGATCT TCATGCGAAC TTCAGTGAAG	720
ATTCATACA TTGGCCTCAT GACCCAGAGC TCCTTGGAGA CACATCACTA TGTGGATTGT	780
GGAGGAAATT CCACAGCTAT TTAACAAC TGCTATTGGTTC TTCCACACAG CGCCTGTAGA	840
AGAGAGCACA GCATATGTTC CCAAGGCCTG AGTTCTGGAC CTACCCCCAC GTGGTGTAAAG	900
CAGAGGAGGA ATTGTTTCAC TTAAC TCCCA GCAAACATCC TCCTGCCACT TAGGAGGAAA	960
CACCTCCCTA TGGTACCATT TATGTTTCTC AGAACCAGCA GAATCAGTGC CTAGCCTGTG	1020
CCCAGCAAAT AGTTGGCACT CAATAAAGAT TTGCAGAATT TAAAAAAAAA AAAAAAAAAA	1080
AAAAAAATTC CTGCGGCCGC	1100

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1588 base pairs
 (B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

GAATTCGGCA	CGAGGGTACC	TGCTTTTCTA	TTGCCTCTTT	GAAACAATGG	TCACGTGTTT	60
CCATGTTCCC	TACTCGGCTC	TCACCATGTT	CATCAGCACC	GAGCAGACTG	AGCGGGATTC	120
TGCCACCGCC	TATCGGATGA	CTGTGGAAGT	GCTGGGCACA	GTGCTGGGCA	CGGCGATCCA	180
GGGACAAATC	GTGGGCCAAG	CAGACACGCC	TTGTTTCCAG	GACCTCAATA	GCTCTACAGT	240
AGCTTCACAA	AGTGCCAACC	ATACACATGG	CACCACCTCA	CACAGGGAAA	CGCAAAAGGC	300
ATACCTGCTG	GCAGCGGGGG	TCATTGTCTG	TATCTATATA	ATCTGTGCTG	TCATCCTGAT	360
CCTGGGCGTG	CGGGAGCAGA	GAGAACCCTA	TGAAGCCCAG	CAGTCTGAGC	CAATCGCCTA	420
CTTCCGGGGC	CTACGGCTGG	TCATGAGCCA	CGGCCCATAC	ATCAAACCTA	TTACTGGCTT	480
CCTCTTCACC	TCCTTGGCTT	TCATGCTGGT	GGAGGGGAAC	TTTGTCTTGT	TTTGCACCTA	540
CACCTTGGGC	TTCCGCAATG	AATTCCAGAA	TCTACTCCTG	GCCATCATGC	TCTCGGCCAC	600
TTTAACCATT	CCCATCTGGC	AGTGGTTCTT	GACCCGGTTT	GGCAAGAAGA	CAGCTGTATA	660
TGTTGGGATC	TCATCAGCAG	TGCCATTTCT	CATCTTGGTG	GCCCTCATGG	AGAGTAACCT	720
CATCATTACA	TATGCGGTAG	CTGTGGCAGC	TGGCATCAGT	GTGGCAGCTG	CCTTCTTACT	780
ACCCTGGTCC	ATGCTGCCTG	ATGTCATTGA	CGACTTCCAT	CTGAAGCAGC	CCCACTTCCA	840
TGGAACCGAG	CCCATCTTCT	TCTCCTTCTA	TGTCTTCTTC	ACCAAGTTTG	CCTCTGGAGT	900
GTCACTGGGC	ATTTCTACCC	TCAGTCTGGA	CTTTGCAGGG	TACCAGACCC	GTGGCTGCTC	960
GCAGCCGGAA	CGTGTCAAGT	TTACACTGAA	CATGCTCGTG	ACCATGGCTC	CCATAGTTCT	1020
CATCCTGCTG	GGCCTGCTGC	TCTTCAAAAT	GTACCCCAT	GATGAGGAGA	GGCGGCGGCA	1080
GAATAAGAAG	GCCCTGCAGG	CACTGAGGGA	CGAGGCCAGC	AGCTCTGGCT	GCTCAGAAAC	1140
AGACTCCACA	GAGCTGGCTA	GCATCCTCTA	GGGCCCGCCA	CGTTGCCCGA	AGCCACCATG	1200
CAGAAGGCCA	CAGAAGGGAT	CAGGACCTGT	CTGCCGGCTT	GCTGAGCAGC	TGGACTGCAG	1260
GTGCTAGGAA	GGGAAGTAA	GACTCAAGGA	GGTGGCCCAG	GACACTTGCT	GTGCTCACTG	1320
TGGGGCCGGC	TGCTCTGTGG	CCTCCTGCCT	CCCCTCTGCC	TGCCTGTGGG	GCCAAGCCCT	1380
GGGGCTGCCA	CTGTGAATAT	GCCAAGGACT	GATCGGGCCT	AGCCCGGAAC	ACTAATGTAG	1440
AAACCTTTTT	TTTACAGAGC	CTAATTAATA	ACTTAATGAC	TGTGTACATA	GCAATGTGTG	1500
TGTATGTATA	TGTCTGTGAG	CTATTAATGT	TATTAATTTT	CATAAAAGCT	GGAAAGCAAA	1560
AAAAAAAAAA	AAAAATTCCT	GCGGCCGC				1588

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1535 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

```

GAATTCGGCA CGAGGCGGAA GTCCCGTCTC ACGGTTGCCC TGGCAGCGCG CGAGGCTGGT      60
GAGTCGGCAG CCCTGTGGCA GCCGGCGGGC TGGTTTCCAT GGTTCACGA TTAGGAACCA      120
CCAGCTGCTG CATCCCATGG CCAGGGGTGG CGTCCAGGTG GCAGAGCAGC TAGGAACGCA      180
AGGCCTGAAC CTGGGGCCAG ACACCCTGCT CTCCCGGCCA TGGTCAACGA CCCTCCAGTA      240
CCTGCCTTAC TGTGGGCCCC GGAGGTGGGC CAAGTCTTGG CAGGCCGTGC CCGCAGGCTG      300
CTGCTGCAGT TTGGGGTGCT CTTCTGCACC ATCCTCCTTT TGCTCTGGGT GTCTGTCTTC      360
CTCTATGGCT CCTTCTACTA TTCCTATATG CCGACAGTCA GCCACCTCAG CCCTGTGCAT      420
TTCTACTACA GGACCGACTG TGATTCCTCC ACCACCTCAC TCTGCTCCTT CCCTGTTGCC      480
AATGTCTCGC TGAATAAGGG TGGACGTGAT CGGGTGCTGA TGTATGGACA GCCGTATCGT      540
GTTACCTTAG AGCTTGAGCT GCCAGAGTCC CCTGTGAATC AAGATTGGG CATGTTCTTG      600
GTCACCATT CTGCTACAC CAGAGGTGGC CGAATCATCT CCACTTCTTC GCGTTCGGTG      660
ATGCTGCATT ACCGCTCAGA CCTGCTCCAG ATGCTGGACA CACTGGTCTT CTCTAGCCTC      720
CTGCTATTTG GCTTTGCAGA GCAGAGGAGG CTGCTGGAGG TGGAACTCTA GGCAGACTAT      780
AGAGAGAACT CGTACGTGCC GACCACTGGA GCGATCATTG AGATCCACAG CAAGCGCATC      840
CAGCTGTATG GAGCCTACCT CCGCATCCAC GCGCACTTCA CTGGGCTCAG ATACCTGCTA      900
TACAACTTCC CGATGACCTG CGCCTTCATA GGTGTTGCCA GCAACTTCAC CTTCTCAGC      960
GTCATCGTGC TCTTCAGCTA CATGCAGTGG GTGTGGGGGG GCATCTGGCC CCGACACCGC     1020
TTCTCTTTGC AGGTTAACAT CCGAAAAAGA GACAATTCCC GGAAGGAAGT CCAACGAAGG     1080
ATCTCTGCTC ATCAGCCAGG GCCTGAAGGC CAGGAGGAGT CAACTCCGCA ATCAGATGTT     1140
ACAGAGGATG GTGAGAGCCC TGAAGATCCC TCAGGGACAG AGGTCAGCTG TCCGAGGAGG     1200
AGAAACCAGA TCAGCAGCCC CTGAGCGGAG AAGAGGAGCT AGAGCCTGAG GCCAGTGATG     1260
G TTCAGGCTC CTGGGAAGAT GCAGCTTTGC TGACGGAGGC CAACCTGCCT GCTCCTGCTC     1320
CTGCTTCTGC TTCTGCCCCCT GTCCTAGAGA CTCTGGGCAG CTCTGAACCT GCTGGGGGTG     1380
CTCTCCGACA GCGCCCCACC TGCTCTAGTT CCTGAAGAAA AGGGGCAGAC TCCTCACATT     1440
CCAGCACTTT CCCACCTGAC TCCTCTCCCC TCGTTTTTCC TTCAATAAAC TATTTTGTGT     1500
CAAAAAAAAA AAAAAAAAAA AATTCCTGCG GCCGC                                     1535

```

(2) INFORMATION FOR SEQ ID NO:16:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1322 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

GAATTCGGCA CGAGGGCGGG CGCTACGGGC TTGACTCCCC CAAGGCCGAG GTCCGCGGCC	60
AGGTGCTGGC GCCGCTGCCC CTCCACGGAG TTGCTGATCA TCTGGGCTGT GATCCACAAA	120
CCCGGTTCTT TGTCCCTCCT AATATCAAAC AGTGGATTGC CTTGCTGCAG AGGGGAAACT	180
GCACGTTTAA AGAGAAAATA TCACGGGCCG CTTTCCACAA TGCAGTTGCT GTAGTCATCT	240
ACAATAATAA ATCCAAAGAG GAGCCAGTTA CCATGACTCA TCCAGGCACT GGAGATATTA	300
TTGCTGTCAT GATAACAGAA TTGAGGGGTA AGGATATTTT GAGTTATCTG GAGAAAAACA	360
TCTCTGTACA AATGACAATA GCTGTTGGAA CTCGAATGCC ACCGAAGAAC TTCAGCCGTG	420
GCTCTCTAGT CTTCTGTGCA ATATCCTTTA TTGTTTTGAT GATTATTTCT TCAGCATGGC	480
TCATATTCTA CTTTATTCAA AAGATCAGGT ACACAAATGC ACGCGACAGG AACCAGCGTC	540
GTCTCGGAGA TGCAGCCAAG AAAGCCATCA GTAAATTGAC AACCAGGACA GTAAAGAAGG	600
GTGACAAGGA AACTGACCCA GACTTTGATC ATTGTGCAGT CTGCATAGAG AGCTATAAGC	660
AGAATGATGT CGTCCGAATT CTCCCCTGCA AGCATGTTTT CCACAAATCC TGCCTGGATC	720
CCTGGCTTAG TGAACATTGT ACCTGTCCTA TGTGCAAACT TAATATATTG AAGGCCCTGG	780
GAATTGTGCC GAATTTGCCA TGTACTGATA ACGTAGCATT CGATATGGAA AGGCTCACCA	840
GAACCCAAGC TGTTAACCGA AGATCAGCCC TCGGCGACCT CGCCGGCGAC AACTCCCTTG	900
GCCTTGAGCC ACTTCGAACT TCGGGGATCT CACCTCTTCC TCAGGATGGG GAGCTCACTC	960
CGAGAACAGG AGAAATCAAC ATTGCAGTAA CAAAAGAATG GTTTATTATT GCCAGTTTGT	1020
GCCTCCTCAG TGCCCTCACA CTCTGCTACA TGATCATCAG AGCCACAGCT AGCTTGAATG	1080
CTAATGAGGT AGAATGGTTT TGAAGAAGAA AAAACCTGCT TTCTGACTGA TTTTGCCTTG	1140
AAGGAAAAAA GAACCTATTT TTGTGCATCA TTTACCAATC ATGCCACACA AGCATTATTT	1200
TTTAGTACAT TTTATTTTTT CATAAAATTG CTAATGCCAA AGCTTTGTAT TAAAAGAAAT	1260
AAATAATAAA ATAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAT TCCTGCGGCC	1320
GC	1322

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1711 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

```

GAATTCGGCA CGAGGCCCTC CCGCGCTCCC GGGGCGCGCG GGCCGCGCCC CCGACGCCCT    60
ACATATACTC AGGTGCGCCC CACCTGTCCG CCCGCACCTG CTGGCTCACC TCCGAGCCAC    120
CTCTGCTGCG CACCGCAGCC TCGGACCTAC AGCCCAGGAT ACTTTGGGAC TTGCCGGCGC    180
TCAGAAACGC GCCCAGACGG CCCCTCCACC TTTTGTTTGC CTAGGGTCGC CGAGAGCGCC    240
CGGAGGGAAC CGCCTGGCCT TCGGGGACCA CCAATTTTGT CTGGAACCAC CCTCCCGGCG    300
TATCCTACTC CCTGTGCCGC GAGGCCATCG CTTCACTGGA GGGGTGCGATT TGTGTGTAGT    360
TTGGTGACAA GATTTCATT CACCTGGCCC AAACCCTTTT TGTCTCTTTG GGTGACCGGA    420
AAACTCCACC TCAAGTTTTT TTTTGTGGGG CTGCCCCCA AGTGTCGTTT GTTTTACTGT    480
AGGGTCTCCC GCCCGGCGCC CCCAGTGTTC TCTGAGGGCG GAAATGGCCA ATTCGGGCCT    540
GCAGTTGCTG GGCTTCTCCA TGGCCCTGCT GGGCTGGGTG GGTCTGGTGG CCTGCACCGC    600
CATCCCGCAG TGGCAGATGA GCTCCTATGC GGGTGACAAC ATCATCACGG CCCAGGCCAT    660
GTACAAGGGG CTGTGGATGG ACTGCGTCAC GCAGAGCACG GGGATGATGA GCTGCAAAAT    720
GTACGACTCG GTGCTCGGCG TGTCGCGGCG CTTGCGAGGCG ACTCGAGGCG TAATGGTGGT    780
CTCCCTGGTG CTGGGCTTCC TGGCCATGTT TGTGGCCACG ATGGGCATGA AGTGCACGCG    840
CTGTGGGGGA GACGACAAAG TGAAGAAGGC CCGTATAGCC ATGGGTGGAG GCATAATTTT    900
CATCGTGGCA GGTCTTGCCG CCTTGGTAGC TTGCTCCTGG TATGGCCATC AGATTGTCAC    960
AGACTTTTAT AACCCTTTGA TCCCTACCAA CATTAAGTAT GAGTTTGGCC CTGCCATCTT   1020
TATTGGCTGG GCAGGGTCTG CCCTAGTCAT CCTGGGAGGT GCACTGCTCT CCTGTTCCCTG   1080
TCCTGGGAAT GAGAGCAAGG CTGGGTACCG TGCACCCCGC TCTTACCCTA AGTCCAATC   1140
TTCCAAGGAG TATGTGTGAC CTGGGATCTC CTTGCCCCAG CCTGACAGGC TATGGGAGTG   1200
TCTAGATGCC TGAAAGGGCC TGGGGCTGAG CTCAGCCTGT GGGCAGGGTG CCGGACAAAG   1260
GCCTCCTGGT CACTCTGTCC CTGCACTCCA TGTATAGTCC TCTTGGGTG GGGGTGGGGG   1320
GGTGGCGTTG GTGGGAGAGA CAAAAGAGG GAGAGTGTGC TTTTGTACA GTAATAAAAA   1380
ATAAGTATTG GGAAGCAGGC TTTTTCCTC TCAGGGCCTC TGCTTTCCTC CCGTCCAGAT   1440
CCTTGCAGGG AGCTTGAAC CTTAGTGCAC CTACTTCAGT TCAGAACACT TAGCACCCCA   1500
CTGACTCCAC TGACAATTGA CTAAAAGATG CAGGTGCTCG TATCTCGACA TTCATTCCCA   1560
CCCCCTCTT ATTTAAATAG CTACCAAAGT ACTTCTTTTT TAATAAAAAA ATAAAGATTT   1620

```

TTATTAGGTA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA 1680
 AAAAAAAAAA AAAAAAATT CCTGCGGCCG C 1711

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1553 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

GAATTCGGCA	CGAGGGCAGG	TCCAGAGTAA	AGTCACTGAA	GAGTGGGAAGC	GAGGAAGGAA	60
CAGGATGATT	AGACCTCAGC	TGCGGACCGC	GGGGCTGGGA	CGATGCCTCC	TGCCGGGGGCT	120
GCTGCTGCTC	CTGGTGCCCG	TCCTCTGGGC	CGGGGCTGAA	AAGCTACATA	CCCAGCCCTC	180
CTGCCCCGCG	GTCTGCCAGC	CCACGCGCTG	CCCCGCGCTG	CCCACCTGCG	CGCTGGGGAC	240
CACGCCGGTG	TTGCACTGT	GCCGCTGTTG	CCGCGTCTGC	CCCGCGGCCG	AGCGTGAAGT	300
CTGCGGCGGG	GCGCAGGGCC	AACCGTGCGC	CCCGGGGCTG	CAGTGCCTCC	AGCCGCTGCG	360
CCCCGGGTTC	CCCAGCACCT	GCGGTTGCCC	GACGCTGGGA	GGGGCCGTGT	GCGGCAGCGA	420
CAGGCGCACC	TACCCAGCA	TGTGCGCGCT	CCGGGCCGAA	AACCGCGCCG	CGCGCCGCCT	480
GGGCAAGGTC	CCGGCCGTGC	CTGTGCAGTG	GGGGAAGTGC	GGGGATACAG	GGACCAGAAG	540
CGCAGGCCCC	CTCAGGAGGA	ATTACAATT	CATCGCCGCG	GTGGTGGAGA	AGGTGGCGCC	600
ATCGGTGGTT	CACGTGCAGC	TGTGGGGCAG	GTTACTTCAC	GGCAGCAGGC	TTGTTCTCTGT	660
GTACAGTGGC	TCTGGGTTCA	TAGTGTCTGA	GGACGGGCTC	ATTATTACCA	ATGCCCATGT	720
TGTCAGGAAC	CAGCAGTGA	TTGAGGTGGT	GCTCCAGAAT	GGGGCCCGTT	ATGAAGCTGT	780
TGTCAAGGAT	ATTGACCTTA	AATTGGATCT	TGCGGTGATT	AAGATTGAAT	CAAATGCTGA	840
ACTTCCTGTA	CTGATGCTGG	GAAGATCATC	TGACCTTCGG	GCTGGAGACT	TTGTGGTGGC	900
TTTGGGCAGC	CCATTTTCTC	TGCAGAACAC	AGCTACTGCA	GGAATTGTCA	GCACCAAACA	960
GCGAGGGGGC	AAAGAACTGG	GGATGAAGGA	TTCAGATATG	GACTACGTCC	AGATTGATGC	1020
CACAATTAAC	TATGGGAATT	CTGGTGGTCC	TCTGGTGAAC	TTGGATGGTG	ATGTGATTGG	1080
CGTCAATTCA	TTGAGGGTGA	CTGATGGAAT	CTCCTTTGCA	ATTCCCTCAG	ATCGAGTTAG	1140
GCAGTTCTTG	GCAGAATACC	ATGAGCACCA	GATGAAAGGA	AAGGCGTTTT	CAAATAAGAA	1200
ATATCTGGGT	CTGCAAATGC	TGTCCCTCAC	TGTGCCCTT	AGTGAAGAAT	TGAAAATGCA	1260
TTATCCAGAT	TTCCCTGATG	TGAGTTCTGG	GGTTTATGTA	TGTAAAGTGG	TTGAAGGAAC	1320

AGCTGCTCAA	AGCTCTGGAT	TGAGAGATCA	CGATGTAATT	GTCAACATAA	ATGGGAAACC	1380
TATTACTACT	ACAACTGATG	TTGTTAAAGC	TCTTGACAGT	GATTCCCTTT	CCATGGCTGT	1440
TCTTCGGGGA	AAAGATAATT	TGCTCCTGAC	AGTCATACCT	GAAACAATCA	ATTAAATATC	1500
TTGTTTTAAA	GTGGGATTAT	CTAAAAAAA	AAAAAAAAA	TTCCTGCGGC	CGC	1553

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1596 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

GAATTCGGCA	CGAGGGGAGC	CGCTCCCGGA	GCCCGGCCGT	AGAGGCTGCA	ATCGCAGCCG	60
GGAGCCCGCA	GCCCGCGCCC	CGAGCCCGCC	GCCGCCCTTC	GAGGGCGCCC	CAGGCCGCGC	120
CATGGTGAAG	GTGACGTTCA	ACTCCGCTCT	GGCCCAAGAG	GAGGCCAAGA	AGGACGAGCC	180
CGAGAGCGGC	GAGGAGGCGC	TCATCATCCC	CCCCGACGCC	GTCGCGGTGG	ACTGCAAGGA	240
CCCAGATGAT	GTGGTACCAG	TTGGCCAAAG	AAGAGCCTGG	TGTTGGTGCA	TGTGCTTTGG	300
ACTAGCAFTT	ATGCTTGCA	GTGTTATTCT	AGGAGGAGCA	TACTTGTACA	AATATTTTGC	360
ACTTCAACCA	GATGACGTGT	ACTACTGTGG	AATAAAGTAC	ATCAAAGATG	ATGTCATCTT	420
AAATGAGCCC	TCTGCAGATG	CCCCAGCTGC	TCTCTACCAG	ACAATTGAAG	AAAATATTAA	480
AATCTTTGAA	GAAGAAGAAG	TTGAATTTAT	CAGTGTGCCT	GTCCCAGAGT	TTGCAGATAG	540
TGATCCTGCC	AACATTGTTC	ATGACTTTAA	CAAGAACTT	ACAGCCTATT	TAGATCTTAA	600
CCTGGATAAG	TGCTATGTGA	TCCCTCTGAA	CACTTCCATT	GTTATGCCAC	CCAGAAACCT	660
ACTGGAGTTA	CTTATTAACA	TCAAGGCTGG	AACCTATTTG	CCTCAGTCCT	ATCTGATTCA	720
TGAGCACATG	GTTATTACTG	ATCGCATTGA	AAACATTGAT	CACCTGGGTT	TCTTTATTTA	780
TCGACTGTGT	CATGACAAGG	AAACTTACAA	ACTGCAACGC	AGAGAAACTA	TTAAAGGTAT	840
TCAGAAACGT	GAAGCCAGCA	ATTGTTTCGC	AATTCGGCAT	TTTGAAAACA	AATTTGCCGT	900
GGAAACTTTA	ATTTGTTCTT	GAACAGTCAA	GAAAAACATT	ATTGAGGAAA	ATTAATATCA	960
CAGCATAACC	CCACCCTTTA	CATTTTGTGC	AGTGATATTT	TTTAAAGTCT	CTTTCATGTA	1020
AGTAGCAAAC	AGGGCTTTAC	TATCTTTTCA	TCTCATTAAT	TCAATTAAAA	CCATTACCTT	1080
AAAATTTTTT	TCTTTCGAAG	TGTGGTGTCT	TTTATATTTG	AATTAGTAAC	TGTATGAAGT	1140

CATAGATAAT AGTACATGTC ACCTTAGGTA GTAGGAAGAA TTACAATTC TTTAAATCAT	1200
TTATCTGGAT TTTTATGTTT TATTAGCATT TTCAAGAAGA CGGATTATCT AGAGAATAAT	1260
CATATATATG CATACGTAAA AATGGACCAC AGTGACTTAT TTGTAGTTGT TAGTTGCCCT	1320
GCTACCTAGT TTGTTAGTGC ATTTGAGCAC ACATTTTAAT TTTCTCTAA TTAATATGTG	1380
CAGTATTTTC AGTGTCAAAT ATATTTAACT ATTTAGAGAA TGATTTCAC CTTTATGTTT	1440
TAATATCCTA GGCATCTGCT GTAATAATAT TTTAGAAAAT GTTTGGAATT TAAGAAATAA	1500
CTTGTGTTAC TAATTTGTAT AACCCATATC TGTGCAATGG AATATAAATA TCACAAAGTT	1560
GTTTAAAAAA AAAAAAAAAA AAATTCCTGC GGCCGC	1596

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 400 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

Met	Ala	Trp	Arg	Arg	Arg	Glu	Ala	Gly	Val	Gly	Ala	Arg	Gly	Val	Leu
1				5					10					15	
Ala	Leu	Ala	Leu	Leu	Ala	Leu	Ala	Leu	Cys	Val	Pro	Gly	Ala	Arg	Gly
			20					25					30		
Arg	Ala	Leu	Glu	Trp	Phe	Ser	Ala	Val	Val	Asn	Ile	Glu	Tyr	Val	Asp
			35				40					45			
Pro	Gln	Thr	Asn	Leu	Thr	Val	Trp	Ser	Val	Ser	Glu	Ser	Gly	Arg	Phe
			50				55				60				
Gly	Asp	Ser	Ser	Pro	Lys	Glu	Gly	Ala	His	Gly	Leu	Val	Gly	Val	Pro
65				70					75				80		
Trp	Ala	Pro	Gly	Gly	Asp	Leu	Glu	Gly	Cys	Ala	Pro	Asp	Thr	Arg	Phe
			85					90					95		
Phe	Val	Pro	Glu	Pro	Gly	Gly	Arg	Gly	Ala	Ala	Pro	Trp	Val	Ala	Leu
			100				105						110		
Val	Ala	Arg	Gly	Gly	Cys	Thr	Phe	Lys	Asp	Lys	Val	Leu	Val	Ala	Ala

115	120	125
Arg Arg Asn Ala Ser Ala Val Val Leu Tyr Asn Glu Glu Arg Tyr Gly		
130	135	140
Asn Ile Thr Leu Pro Met Ser His Ala Gly Thr Gly Asn Ile Val Val		
145	150	155
Ile Met Ile Ser Tyr Pro Lys Gly Arg Glu Ile Leu Glu Leu Val Gln		
165	170	175
Lys Gly Ile Pro Val Thr Met Thr Ile Gly Val Gly Thr Arg His Val		
180	185	190
Gln Glu Phe Ile Ser Gly Gln Ser Val Val Phe Val Ala Ile Ala Phe		
195	200	205
Ile Thr Met Met Ile Ile Ser Leu Ala Trp Leu Ile Phe Tyr Tyr Ile		
210	215	220
Gln Arg Phe Leu Tyr Thr Gly Ser Gln Ile Gly Ser Gln Ser His Arg		
225	230	235
Lys Glu Thr Lys Lys Val Ile Gly Gln Leu Leu Leu His Thr Val Lys		
245	250	255
His Gly Glu Lys Gly Ile Asp Val Asp Ala Glu Asn Cys Ala Val Cys		
260	265	270
Ile Glu Asn Phe Lys Val Lys Asp Ile Ile Arg Ile Leu Pro Cys Lys		
275	280	285
His Ile Phe His Arg Ile Cys Ile Asp Pro Trp Leu Leu Asp His Arg		
290	295	300
Thr Cys Pro Met Cys Lys Leu Asp Val Ile Lys Ala Leu Gly Tyr Trp		
305	310	315
Gly Glu Pro Gly Asp Val Gln Glu Met Pro Ala Pro Glu Ser Pro Pro		
325	330	335
Gly Arg Asp Pro Ala Ala Asn Leu Ser Leu Ala Leu Pro Asp Asp Asp		
340	345	350
Gly Ser Asp Asp Ser Ser Pro Pro Ser Ala Ser Pro Ala Glu Ser Glu		
355	360	365
Pro Gln Cys Asp Pro Ser Phe Lys Gly Asp Ala Gly Glu Asn Thr Ala		
370	375	380
Leu Leu Glu Ala Gly Arg Ser Asp Ser Arg His Gly Gly Pro Ile Ser		
385	390	395
		400

(2) INFORMATION FOR SEQ ID NO:21:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 291 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

```

Met Asp Lys Gly Ser Ala Gly His Pro Gly Gly Val Leu Val Trp Gly
 1             5             10             15
Arg Ser Pro Ala Pro Thr Ala Leu Trp Gly Ala Ser Pro Trp Leu Ser
      20             25             30
Pro Leu Thr Ser Ala Leu Arg Gln Pro Leu His Arg Ala Pro Leu Leu
      35             40             45
Pro Gly Gln Leu Cys Trp Ser Pro Arg Pro Leu Glu Lys Asn Lys Ala
      50             55             60
Met Gly Arg Pro Leu Leu Leu Pro Leu Leu Leu Leu Leu Gln Pro Pro
65             70             75             80
Ala Phe Leu Gln Pro Gly Gly Ser Thr Gly Ser Gly Pro Ser Tyr Leu
      85             90             95
Tyr Gly Val Thr Gln Pro Lys His Leu Ser Ala Ser Met Gly Gly Ser
      100            105            110
Val Glu Ile Pro Phe Ser Phe Tyr Tyr Pro Trp Glu Leu Ala Ile Val
      115            120            125
Pro Asn Val Arg Ile Ser Trp Arg Arg Gly His Phe His Gly Gln Ser
      130            135            140
Phe Tyr Ser Thr Arg Pro Pro Ser Ile His Lys Asp Tyr Val Asn Arg
145            150            155            160
Leu Phe Leu Asn Trp Thr Glu Gly Gln Glu Ser Gly Phe Leu Arg Ile
      165            170            175
Ser Asn Leu Arg Lys Glu Asp Gln Ser Val Tyr Phe Cys Arg Val Glu
      180            185            190

```

Leu Asp Thr Arg Arg Ser Gly Arg Gln Gln Leu Gln Ser Ile Lys Gly
 195 200 205
 Thr Lys Leu Thr Ile Thr Gln Ala Val Thr Thr Thr Thr Thr Trp Arg
 210 215 220
 Pro Ser Ser Thr Thr Thr Ile Ala Gly Leu Arg Val Thr Glu Ser Lys
 225 230 235 240
 Gly His Ser Glu Ser Trp His Leu Ser Leu Asp Thr Ala Ile Arg Val
 245 250 255
 Ala Leu Ala Val Ala Val Leu Lys Thr Val Ile Leu Gly Leu Leu Cys
 260 265 270
 Leu Leu Leu Leu Trp Trp Arg Arg Arg Lys Gly Ser Arg Ala Pro Ser
 275 280 285
 Ser Asp Phe
 290

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 293 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

Met Thr Val Ser Gln Arg Phe Gln Leu Ser Asn Ser Gly Pro Asn Ser
 1 5 10 15
 Thr Ile Lys Met Lys Ile Ala Leu Arg Val Leu His Leu Glu Lys Arg
 20 25 30
 Glu Arg Pro Pro Asp His Gln His Ser Ala Gln Val Lys Arg Pro Ser
 35 40 45
 Val Ser Lys Glu Gly Arg Lys Thr Ser Ile Lys Ser His Met Ser Gly
 50 55 60
 Ser Pro Gly Pro Gly Gly Ser Asn Thr Ala Pro Ser Thr Pro Val Ile

65		70		75		80
Gly Gly Ser Asp Lys Pro Gly Met Glu Glu Lys Ala Gln Pro Pro Glu						
	85		90		95	
Ala Gly Pro Gln Gly Leu His Asp Leu Gly Arg Ser Ser Ser Ser Leu						
	100		105		110	
Leu Ala Ser Pro Gly His Ile Ser Val Lys Glu Pro Thr Pro Ser Ile						
	115		120		125	
Ala Ser Asp Ile Ser Leu Pro Ile Ala Thr Gln Glu Leu Arg Gln Arg						
	130		135		140	
Leu Arg Gln Leu Glu Asn Gly Thr Thr Leu Gly Gln Ser Pro Leu Gly						
	145		150		155	160
Gln Ile Gln Leu Thr Ile Arg His Ser Ser Gln Arg Asn Lys Leu Ile						
	165		170		175	
Val Val Val His Ala Cys Arg Asn Leu Ile Ala Phe Ser Glu Asp Gly						
	180		185		190	
Ser Asp Pro Tyr Val Arg Met Tyr Leu Leu Pro Asp Lys Arg Arg Ser						
	195		200		205	
Gly Arg Arg Lys Thr His Val Ser Lys Lys Thr Leu Asn Pro Val Phe						
	210		215		220	
Asp Gln Ser Phe Asp Phe Ser Val Ser Leu Pro Glu Val Gln Arg Arg						
	225		230		235	240
Thr Leu Asp Val Ala Val Lys Asn Ser Gly Gly Phe Leu Ser Lys Asp						
	245		250		255	
Lys Gly Leu Leu Gly Lys Val Leu Val Ala Leu Ala Ser Glu Glu Leu						
	260		265		270	
Ala Lys Gly Trp Thr Gln Trp Tyr Asp Leu Thr Glu Asp Gly Thr Arg						
	275		280		285	
Pro Gln Ala Met Thr						
	290					

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 206 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

Met	Glu	Arg	Arg	His	Pro	Val	Cys	Ser	Gly	Thr	Cys	Gln	Pro	Thr	Gln
1				5					10					15	
Phe	Arg	Cys	Ser	Asn	Gly	Cys	Cys	Ile	Asp	Ser	Phe	Leu	Glu	Cys	Asp
			20					25					30		
Asp	Thr	Pro	Asn	Cys	Pro	Asp	Ala	Ser	Asp	Glu	Ala	Ala	Cys	Glu	Lys
			35				40					45			
Tyr	Thr	Ser	Gly	Phe	Asp	Glu	Leu	Gln	Arg	Ile	His	Phe	Pro	Ser	Asp
	50						55				60				
Lys	Gly	His	Cys	Val	Asp	Leu	Pro	Asp	Thr	Gly	Leu	Cys	Lys	Glu	Ser
65					70					75				80	
Ile	Pro	Arg	Trp	Tyr	Tyr	Asn	Pro	Phe	Ser	Glu	His	Cys	Ala	Arg	Phe
				85					90					95	
Thr	Tyr	Gly	Gly	Cys	Tyr	Gly	Asn	Lys	Asn	Asn	Phe	Glu	Glu	Glu	Gln
			100					105					110		
Gln	Cys	Leu	Glu	Ser	Cys	Arg	Gly	Ile	Ser	Lys	Lys	Asp	Val	Phe	Gly
			115					120					125		
Leu	Arg	Arg	Glu	Ile	Pro	Ile	Pro	Ser	Thr	Gly	Ser	Val	Glu	Met	Ala
			130				135					140			
Val	Ala	Val	Phe	Leu	Val	Ile	Cys	Ile	Val	Val	Val	Val	Ala	Ile	Leu
145					150					155				160	
Gly	Tyr	Cys	Phe	Phe	Lys	Asn	Gln	Arg	Lys	Asp	Phe	His	Gly	His	His
				165					170				175		
His	His	Pro	Pro	Pro	Thr	Pro	Ala	Ser	Ser	Thr	Val	Ser	Thr	Thr	Glu
				180				185					190		
Asp	Thr	Glu	His	Leu	Val	Tyr	Asn	His	Thr	Thr	Arg	Pro	Leu		
			195					200					205		

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 220 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

```

Met Ala Gly Leu Ser Arg Gly Ser Ala Arg Ala Leu Leu Ala Ala Leu
 1             5             10             15
Leu Ala Ser Thr Leu Leu Ala Leu Leu Val Ser Pro Ala Arg Gly Arg
      20             25             30
Gly Gly Arg Asp His Gly Asp Trp Asp Glu Ala Ser Arg Leu Pro Pro
      35             40             45
Leu Pro Pro Arg Glu Asp Ala Ala Arg Val Ala Arg Phe Val Thr His
      50             55             60
Val Ser Asp Trp Gly Ala Leu Ala Thr Ile Ser Thr Leu Glu Ala Val
      65             70             75             80
Arg Gly Arg Pro Phe Ala Asp Val Leu Ser Leu Ser Asp Gly Pro Pro
      85             90             95
Gly Ala Gly Ser Gly Val Pro Tyr Phe Tyr Leu Ser Pro Leu Gln Leu
      100            105            110
Ser Val Ser Asn Leu Gln Glu Asn Pro Tyr Ala Thr Leu Thr Met Thr
      115            120            125
Leu Ala Gln Thr Asn Phe Cys Lys Lys His Gly Phe Asp Pro Gln Ser
      130            135            140
Pro Leu Cys Val His Ile Met Leu Ser Gly Thr Val Thr Lys Val Asn
      145            150            155            160
Glu Thr Glu Met Asp Ile Ala Lys His Ser Leu Phe Ile Arg His Pro
      165            170            175
Glu Met Lys Thr Trp Pro Ser Ser His Asn Trp Phe Phe Ala Lys Leu
      180            185            190
Asn Ile Thr Asn Ile Trp Val Leu Asp Tyr Phe Gly Gly Pro Lys Ile
      195            200            205
Val Thr Pro Glu Glu Tyr Tyr Asn Val Thr Val Gln

```

210

215

220

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 197 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

```

Met Asp His His Cys Pro Trp Leu Asn Asn Cys Val Gly His Tyr Asn
 1             5             10             15
His Arg Tyr Phe Phe Ser Phe Cys Phe Phe Met Thr Leu Gly Cys Val
      20             25             30
Tyr Cys Ser Tyr Gly Ser Trp Asp Leu Phe Arg Glu Ala Tyr Ala Ala
      35             40             45
Ile Glu Lys Met Lys Gln Leu Asp Lys Asn Lys Leu Gln Ala Val Ala
50 55 60
Asn Gln Thr Tyr His Gln Thr Pro Pro Pro Thr Phe Ser Phe Arg Glu
65             70             75             80
Arg Met Thr His Lys Ser Leu Val Tyr Leu Trp Phe Leu Cys Ser Ser
      85             90             95
Val Ala Leu Ala Leu Gly Ala Leu Thr Val Trp His Ala Val Leu Ile
      100            105            110
Ser Arg Gly Glu Thr Ser Ile Glu Arg His Ile Asn Lys Lys Glu Arg
      115            120            125
Arg Arg Leu Gln Ala Lys Gly Arg Val Phe Arg Asn Pro Tyr Asn Tyr
      130            135            140
Gly Cys Leu Asp Asn Trp Lys Val Phe Leu Gly Val Asp Thr Gly Arg
      145            150            155            160
His Trp Leu Thr Arg Val Leu Leu Pro Ser Thr His Leu Pro His Gly
      165            170            175

```


Asn Gly Met Ser Trp Glu Pro Pro Pro Trp Val Thr Ala His Ser Ala
 180 185 190
 Ser Val Met Ala Val
 195

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 451 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

Met Ala Pro Leu Gly Met Leu Leu Gly Leu Leu Met Ala Ala Cys Phe
 1 5 10 15
 Thr Phe Cys Leu Ser His Gln Asn Leu Lys Glu Phe Ala Leu Thr Asn
 20 25 30
 Pro Glu Lys Ser Ser Thr Lys Glu Thr Glu Arg Lys Glu Thr Lys Ala
 35 40 45
 Glu Glu Glu Leu Asp Ala Glu Val Leu Glu Val Phe His Pro Thr His
 50 55 60
 Glu Trp Gln Ala Leu Gln Pro Gly Gln Ala Val Pro Ala Gly Ser His
 65 70 75 80
 Val Arg Leu Asn Leu Gln Thr Gly Glu Arg Glu Ala Lys Leu Gln Tyr
 85 90 95
 Glu Asp Lys Phe Arg Asn Asn Leu Lys Gly Lys Arg Leu Asp Ile Asn
 100 105 110
 Thr Asn Thr Tyr Thr Ser Gln Asp Leu Lys Ser Ala Leu Ala Lys Phe
 115 120 125
 Lys Glu Gly Ala Glu Met Glu Ser Ser Lys Glu Asp Lys Ala Arg Gln
 130 135 140
 Ala Glu Val Lys Arg Leu Phe Arg Pro Ile Glu Glu Leu Lys Lys Asp

145	150	155	160
Phe Asp Glu Leu Asn Val Val Ile Glu Thr Asp Met Gln Ile Met Val			
	165	170	175
Arg Leu Ile Asn Lys Phe Asn Ser Ser Ser Ser Ser Leu Glu Glu Lys			
	180	185	190
Ile Ala Ala Leu Phe Asp Leu Glu Tyr Tyr Val His Gln Met Asp Asn			
	195	200	205
Ala Gln Asp Leu Leu Ser Phe Gly Gly Leu Gln Val Val Ile Asn Gly			
	210	215	220
Leu Asn Ser Thr Glu Pro Leu Val Lys Glu Tyr Ala Ala Phe Val Leu			
225	230	235	240
Gly Ala Ala Phe Ser Ser Asn Pro Lys Val Gln Val Glu Ala Ile Glu			
	245	250	255
Gly Gly Ala Leu Gln Lys Leu Leu Val Ile Leu Ala Thr Glu Gln Pro			
	260	265	270
Leu Thr Ala Lys Lys Lys Val Leu Phe Ala Leu Cys Ser Leu Leu Arg			
	275	280	285
His Phe Pro Tyr Ala Gln Arg Gln Phe Leu Lys Leu Gly Gly Leu Gln			
	290	295	300
Val Leu Arg Thr Leu Val Gln Glu Lys Gly Thr Glu Val Leu Ala Val			
305	310	315	320
Arg Val Val Thr Leu Leu Tyr Asp Leu Val Thr Glu Lys Met Phe Ala			
	325	330	335
Glu Glu Glu Ala Glu Leu Thr Gln Glu Met Ser Pro Glu Lys Leu Gln			
	340	345	350
Gln Tyr Arg Gln Val His Leu Leu Pro Gly Leu Trp Glu Gln Gly Trp			
	355	360	365
Cys Glu Ile Thr Ala His Leu Leu Ala Leu Pro Glu His Asp Ala Arg			
	370	375	380
Glu Lys Val Leu Gln Thr Leu Gly Val Leu Leu Thr Thr Cys Arg Asp			
385	390	395	400
Arg Tyr Arg Gln Asp Pro Gln Leu Gly Arg Thr Leu Ala Ser Leu Gln			
	405	410	415
Ala Glu Tyr Gln Val Leu Ala Ser Leu Glu Leu Gln Asp Gly Glu Asp			
	420	425	430
Glu Gly Tyr Phe Gln Glu Leu Leu Gly Ser Val Asn Ser Leu Leu Lys			

435 440 445
 Glu Leu Arg
 450

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 254 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

Met	Trp	Gln	Ala	Gly	Lys	Arg	Gln	Ala	Ser	Arg	Ala	Phe	Ser	Leu	Tyr
1				5					10					15	
Ala	Asn	Ile	Asp	Ile	Leu	Arg	Pro	Tyr	Phe	Asp	Val	Glu	Pro	Ala	Gln
			20					25					30		
Val	Arg	Ser	Arg	Leu	Leu	Glu	Ser	Met	Ile	Pro	Ile	Lys	Met	Val	Asn
			35					40					45		
Phe	Pro	Gln	Lys	Ile	Ala	Gly	Glu	Leu	Tyr	Gly	Pro	Leu	Met	Leu	Val
			50				55						60		
Phe	Thr	Leu	Val	Ala	Ile	Leu	Leu	His	Gly	Met	Lys	Thr	Ser	Asp	Thr
65					70					75					80
Ile	Ile	Arg	Glu	Gly	Thr	Leu	Met	Gly	Thr	Ala	Ile	Gly	Thr	Cys	Phe
					85					90					95
Gly	Tyr	Trp	Leu	Gly	Val	Ser	Ser	Phe	Ile	Tyr	Phe	Leu	Ala	Tyr	Leu
			100						105					110	
Cys	Asn	Ala	Gln	Ile	Thr	Met	Leu	Gln	Met	Leu	Ala	Leu	Leu	Gly	Tyr
			115						120					125	
Gly	Leu	Phe	Gly	His	Cys	Ile	Val	Leu	Phe	Ile	Thr	Tyr	Asn	Ile	His
			130						135					140	
Leu	His	Ala	Leu	Phe	Tyr	Leu	Phe	Trp	Leu	Leu	Val	Gly	Gly	Leu	Ser
145							150						155		160

Thr Leu Arg Met Val Ala Val Leu Val Ser Arg Thr Val Gly Pro Thr
 165 170 175
 Gln Arg Leu Leu Leu Cys Gly Thr Leu Ala Ala Leu His Met Leu Phe
 180 185 190
 Leu Leu Tyr Leu His Phe Ala Tyr His Lys Val Val Glu Gly Ile Leu
 195 200 205
 Asp Thr Leu Glu Gly Pro Asn Ile Pro Pro Ile Gln Arg Val Pro Arg
 210 215 220
 Asp Ile Pro Ala Met Leu Pro Ala Ala Arg Leu Pro Thr Thr Val Leu
 225 230 235 240
 Asn Ala Thr Ala Lys Ala Val Ala Val Thr Leu Gln Ser His
 245 250

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 221 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

Met Gly Ser Glu Asn Glu Ala Leu Asp Leu Ser Met Lys Ser Val Pro
 1 5 10 15
 Trp Leu Lys Ala Gly Glu Val Ser Pro Pro Ile Phe Gln Glu Asp Ala
 20 25 30
 Ala Leu Asp Leu Ser Val Ala Ala His Arg Lys Ser Glu Pro Pro Pro
 35 40 45
 Glu Thr Leu Tyr Asp Ser Gly Ala Ser Val Asp Ser Ser Gly His Thr
 50 55 60
 Val Met Glu Lys Leu Pro Ser Gly Met Glu Ile Ser Phe Ala Pro Ala
 65 70 75 80
 Thr Ser His Glu Ala Pro Ala Met Met Asp Ser His Ile Ser Ser Ser

	85		90		95										
Asp	Ala	Ala	Thr	Glu	Met	Leu	Ser	Gln	Pro	Asn	His	Pro	Ser	Gly	Glu
	100							105						110	
Val	Lys	Ala	Glu	Asn	Asn	Ile	Glu	Met	Val	Gly	Glu	Ser	Gln	Ala	Ala
	115						120						125		
Lys	Val	Ile	Val	Ser	Val	Glu	Asp	Ala	Val	Pro	Thr	Ile	Phe	Cys	Gly
	130					135						140			
Lys	Ile	Lys	Gly	Leu	Ser	Gly	Val	Ser	Thr	Lys	Asn	Phe	Ser	Phe	Lys
145				150						155					160
Arg	Glu	Asp	Ser	Val	Leu	Gln	Gly	Tyr	Asp	Ile	Asn	Ser	Gln	Gly	Glu
				165						170				175	
Glu	Ser	Met	Gly	Asn	Ala	Glu	Pro	Leu	Arg	Lys	Pro	Ile	Lys	Asn	Arg
	180								185				190		
Ser	Ile	Lys	Leu	Lys	Lys	Val	Asn	Ser	Gln	Glu	Val	His	Met	Leu	Pro
	195						200						205		
Ile	Lys	Lys	Gln	Arg	Leu	Ala	Thr	Phe	Phe	Pro	Arg	Lys			
	210						215					220			

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 266 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

Met	Val	Lys	Val	Thr	Phe	Asn	Ser	Ala	Leu	Ala	Gln	Lys	Glu	Ala	Lys
1				5					10					15	
Lys	Asp	Glu	Pro	Lys	Ser	Gly	Glu	Glu	Ala	Leu	Ile	Ile	Pro	Pro	Asp
			20					25					30		
Ala	Val	Ala	Val	Asp	Cys	Lys	Asp	Pro	Asp	Asp	Val	Val	Pro	Val	Gly
			35					40					45		

Gln	Arg	Arg	Ala	Trp	Cys	Trp	Cys	Met	Cys	Phe	Gly	Leu	Ala	Phe	Met
50						55						60			
Leu	Ala	Gly	Val	Ile	Leu	Gly	Gly	Ala	Tyr	Leu	Tyr	Lys	Tyr	Phe	Ala
65					70					75				80	
Leu	Gln	Pro	Asp	Asp	Val	Tyr	Tyr	Cys	Gly	Ile	Lys	Tyr	Ile	Lys	Asp
				85					90				95		
Asp	Val	Ile	Leu	Asn	Glu	Pro	Ser	Ala	Asp	Ala	Pro	Ala	Ala	Leu	Tyr
	100							105					110		
Gln	Thr	Ile	Glu	Glu	Asn	Ile	Lys	Ile	Phe	Glu	Glu	Glu	Glu	Val	Glu
	115						120					125			
Phe	Ile	Ser	Val	Pro	Val	Pro	Glu	Phe	Ala	Asp	Ser	Asp	Pro	Ala	Asn
	130					135					140				
Ile	Val	His	Asp	Phe	Asn	Lys	Lys	Leu	Thr	Ala	Tyr	Leu	Asp	Leu	Asn
145					150					155				160	
Leu	Asp	Lys	Cys	Tyr	Val	Ile	Pro	Leu	Asn	Thr	Ser	Ile	Val	Met	Pro
			165					170					175		
Pro	Arg	Asn	Leu	Leu	Glu	Leu	Leu	Ile	Asn	Ile	Lys	Ala	Gly	Thr	Tyr
	180							185					190		
Leu	Pro	Gln	Ser	Tyr	Leu	Ile	His	Glu	His	Met	Val	Ile	Thr	Asp	Arg
	195						200						205		
Ile	Glu	Asn	Ile	Asp	His	Leu	Gly	Phe	Phe	Ile	Tyr	Arg	Leu	Cys	His
210							215						220		
Asp	Lys	Glu	Thr	Tyr	Lys	Leu	Gln	Arg	Arg	Glu	Thr	Ile	Lys	Gly	Ile
225					230					235				240	
Gln	Lys	Arg	Glu	Ala	Ser	Asn	Cys	Phe	Ala	Ile	Arg	His	Phe	Glu	Asn
			245					250					255		
Lys	Phe	Ala	Val	Glu	Thr	Leu	Ile	Cys	Ser						
	260						265								

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 251 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

```

Met Pro Thr Gly Asp Phe Asp Ser Lys Pro Ser Trp Ala Asp Gln Val
 1           5           10           15
Glu Glu Glu Gly Glu Asp Asp Lys Cys Val Thr Ser Glu Leu Leu Lys
          20           25           30
Gly Ile Pro Leu Ala Thr Gly Asp Thr Ser Pro Glu Pro Glu Leu Leu
          35           40           45
Pro Gly Ala Pro Leu Pro Pro Pro Lys Glu Val Ile Asn Gly Asn Ile
          50           55           60
Lys Thr Val Thr Glu Tyr Lys Ile Asp Glu Asp Gly Lys Lys Phe Lys
65           70           75           80
Ile Val Arg Thr Phe Arg Ile Glu Thr Arg Lys Ala Ser Lys Ala Val
          85           90           95
Ala Arg Arg Lys Asn Trp Lys Lys Phe Gly Asn Ser Glu Phe Asp Pro
          100          105          110
Pro Gly Pro Asn Val Ala Thr Thr Thr Val Ser Asp Asp Val Ser Met
          115          120          125
Thr Phe Ile Thr Ser Lys Glu Asp Leu Asn Cys Gln Glu Glu Glu Asp
          130          135          140
Pro Met Asn Lys Phe Lys Gly Gln Lys Ile Val Ser Cys Arg Ile Cys
145          150          155          160
Lys Gly Asp His Trp Thr Thr Arg Cys Pro Tyr Lys Asp Thr Leu Gly
          165          170          175
Pro Met Gln Lys Glu Leu Ala Glu Gln Leu Gly Leu Ser Thr Gly Glu
          180          185          190
Lys Glu Lys Leu Pro Gly Glu Leu Glu Pro Val Gln Ala Thr Gln Asn
          195          200          205
Lys Thr Gly Lys Tyr Val Pro Pro Ser Leu Arg Asp Gly Ala Ser Arg
          210          215          220
Arg Gly Glu Ser Met Gln Pro Asn Arg Arg Ala Asp Asp Asn Ala Thr
225          230          235          240
Ile Arg Val Thr Asn Leu Arg Arg Gly His Ala
          245          250

```

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 377 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

```

Met Arg Arg Leu Asn Arg Lys Lys Thr Leu Ser Leu Val Lys Glu Leu
 1             5             10             15
Asp Ala Phe Pro Lys Val Pro Glu Ser Tyr Val Glu Thr Ser Ala Ser
      20             25             30
Gly Gly Thr Val Ser Leu Ile Ala Phe Thr Thr Met Ala Leu Leu Thr
      35             40             45
Ile Met Glu Phe Ser Val Tyr Gln Asp Thr Trp Met Lys Tyr Glu Tyr
      50             55             60
Glu Val Asp Lys Asp Phe Ser Ser Lys Leu Arg Ile Asn Ile Asp Ile
65-----70-----75-----80
Thr Val Ala Met Lys Cys Gln Tyr Val Gly Ala Asp Val Leu Asp Leu
      85             90             95
Ala Glu Thr Met Val Ala Ser Ala Asp Gly Leu Val Tyr Glu Pro Thr
      100            105            110
Val Phe Asp Leu Ser Pro Gln Gln Lys Glu Trp Gln Arg Met Leu Gln
      115            120            125
Leu Ile Gln Ser Arg Leu Gln Glu Glu His Ser Leu Gln Asp Val Ile
      130            135            140
Phe Lys Ser Ala Phe Lys Ser Thr Ser Thr Ala Leu Pro Pro Arg Glu
      145            150            155            160
Asp Asp Ser Ser Gln Ser Pro Asn Ala Cys Arg Ile His Gly His Leu
      165            170            175
Tyr Val Asn Lys Val Ala Gly Asn Phe His Ile Thr Val Gly Lys Ala
      180            185            190

```



```

Ile Pro His Pro Arg Gly His Ala His Leu Ala Ala Leu Val Asn His
      195                      200                      205
Glu Ser Tyr Asn Phe Ser His Arg Ile Asp His Leu Ser Phe Gly Glu
      210                      215                      220
Leu Val Pro Ala Ile Ile Asn Pro Leu Asp Gly Thr Glu Lys Ile Ala
      225                      230                      235                      240
Ile Asp His Asn Gln Met Phe Gln Tyr Phe Ile Thr Val Val Pro Thr
      245                      250                      255
Lys Leu His Thr Tyr Lys Ile Ser Ala Asp Thr His Gln Phe Ser Val
      260                      265                      270
Thr Glu Arg Glu Arg Ile Ile Asn His Ala Ala Gly Ser His Gly Val
      275                      280                      285
Ser Gly Ile Phe Met Lys Tyr Asp Leu Ser Ser Leu Met Val Thr Val
      290                      295                      300
Thr Glu Glu His Met Pro Phe Trp Gln Phe Phe Val Arg Leu Cys Gly
      305                      310                      315                      320
Ile Val Gly Gly Ile Phe Ser Thr Thr Gly Met Leu His Gly Ile Gly
      325                      330                      335
Lys Phe Ile Val Glu Ile Ile Cys Cys Arg Phe Arg Leu Gly Ser Tyr
      340                      345                      350
Lys Pro Val Asn Ser Val Pro Phe Glu Asp Gly His Thr Asp Asn His
      355                      360                      365
Leu Pro Leu Leu Glu Asn Asn Thr His
      370                      375

```

(2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 250 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

Met Gly Ser Gln His Ser Ala Ala Ala Arg Pro Ser Ser Cys Arg Arg
 1 5 10 15
 Lys Gln Glu Asp Asp Arg Asp Gly Leu Leu Ala Glu Arg Glu Gln Glu
 20 25 30
 Glu Ala Ile Ala Gln Phe Pro Tyr Val Glu Phe Thr Gly Arg Asp Ser
 35 40 45
 Ile Thr Cys Leu Thr Cys Gln Gly Thr Gly Tyr Ile Pro Thr Glu Gln
 50 55 60
 Val Asn Glu Leu Val Ala Leu Ile Pro His Ser Asp Gln Arg Leu Arg
 65 70 75 80
 Pro Gln Arg Thr Lys Gln Tyr Val Leu Leu Ser Ile Leu Leu Cys Leu
 85 90 95
 Leu Ala Ser Gly Leu Val Val Phe Phe Leu Phe Pro His Ser Val Leu
 100 105 110
 Val Asp Asp Asp Gly Ile Lys Val Val Lys Val Thr Phe Asn Lys Gln
 115 120 125
 Asp Ser Leu Val Ile Leu Thr Ile Met Ala Thr Leu Lys Ile Arg Asn
 130 135 140
 Ser Asn Phe Tyr Thr Val Ala Val Thr Ser Leu Ser Ser Gln Ile Gln
 145 150 155 160
 Tyr Met Asn Thr Val Val Ser Thr Tyr Val Thr Thr Asn Val Ser Leu
 165 170 175
 Ile Pro Pro Arg Ser Glu Gln Leu Val Asn Phe Thr Gly Lys Ala Glu
 180 185 190
 Met Gly Gly Pro Phe Ser Tyr Val Tyr Phe Phe Cys Thr Val Pro Glu
 195 200 205
 Ile Leu Val His Asn Ile Val Ile Phe Met Arg Thr Ser Val Lys Ile
 210 215 220
 Ser Tyr Ile Gly Leu Met Thr Gln Ser Ser Leu Glu Thr His His Tyr
 225 230 235 240
 Val Asp Cys Gly Gly Asn Ser Thr Ala Ile
 245 250

(2) INFORMATION FOR SEQ ID NO:33:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 374 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

```

Met Val Thr Cys Phe His Val Pro Tyr Ser Ala Leu Thr Met Phe Ile
 1             5             10             15
Ser Thr Glu Gln Thr Glu Arg Asp Ser Ala Thr Ala Tyr Arg Met Thr
      20             25             30
Val Glu Val Leu Gly Thr Val Leu Gly Thr Ala Ile Gln Gly Gln Ile
      35             40             45
Val Gly Gln Ala Asp Thr Pro Cys Phe Gln Asp Leu Asn Ser Ser Thr
      50             55             60
Val Ala Ser Gln Ser Ala Asn His Thr His Gly Thr Thr Ser His Arg
      65             70             75             80
Glu Thr Gln Lys Ala Tyr Leu Leu Ala Ala Gly Val Ile Val Cys Ile
      85             90             95
Tyr Ile Ile Cys Ala Val Ile Leu Ile Leu Gly Val Arg Glu Gln Arg
      100            105            110
Glu Pro Tyr Glu Ala Gln Gln Ser Glu Pro Ile Ala Tyr Phe Arg Gly
      115            120            125
Leu Arg Leu Val Met Ser His Gly Pro Tyr Ile Lys Leu Ile Thr Gly
      130            135            140
Phe Leu Phe Thr Ser Leu Ala Phe Met Leu Val Glu Gly Asn Phe Val
      145            150            155            160
Leu Phe Cys Thr Tyr Thr Leu Gly Phe Arg Asn Glu Phe Gln Asn Leu
      165            170            175
Leu Leu Ala Ile Met Leu Ser Ala Thr Leu Thr Ile Pro Ile Trp Gln
      180            185            190
Trp Phe Leu Thr Arg Phe Gly Lys Lys Thr Ala Val Tyr Val Gly Ile
      195            200            205
Ser Ser Ala Val Pro Phe Leu Ile Leu Val Ala Leu Met Glu Ser Asn

```

```

      210              215              220
Leu Ile Ile Thr Tyr Ala Val Ala Val Ala Ala Gly Ile Ser Val Ala
225              230              235              240
Ala Ala Phe Leu Leu Pro Trp Ser Met Leu Pro Asp Val Ile Asp Asp
      245              250              255
Phe His Leu Lys Gln Pro His Phe His Gly Thr Glu Pro Ile Phe Phe
      260              265              270
Ser Phe Tyr Val Phe Phe Thr Lys Phe Ala Ser Gly Val Ser Leu Gly
      275              280              285
Ile Ser Thr Leu Ser Leu Asp Phe Ala Gly Tyr Gln Thr Arg Gly Cys
      290              295              300
Ser Gln Pro Glu Arg Val Lys Phe Thr Leu Asn Met Leu Val Thr Met
305              310              315              320
Ala Pro Ile Val Leu Ile Leu Leu Gly Leu Leu Leu Phe Lys Met Tyr
      325              330              335
Pro Ile Asp Glu Glu Arg Arg Arg Gln Asn Lys Lys Ala Leu Gln Ala
      340              345              350
Leu Arg Asp Glu Ala Ser Ser Ser Gly Cys Ser Glu Thr Asp Ser Thr
      355              360              365
Glu Leu Ala Ser Ile Leu
      370

```

(2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 334 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

```

Met Val Asn Asp Pro Pro Val Pro Ala Leu Leu Trp Ala Gln Glu Val
 1              5              10              15

```

Gly Gln Val Leu Ala Gly Arg Ala Arg Arg Leu Leu Leu Gln Phe Gly
 20 25 30
 Val Leu Phe Cys Thr Ile Leu Leu Leu Leu Trp Val Ser Val Phe Leu
 35 40 45
 Tyr Gly Ser Phe Tyr Tyr Ser Tyr Met Pro Thr Val Ser His Leu Ser
 50 55 60
 Pro Val His Phe Tyr Tyr Arg Thr Asp Cys Asp Ser Ser Thr Thr Ser
 65 70 75 80
 Leu Cys Ser Phe Pro Val Ala Asn Val Ser Leu Thr Lys Gly Gly Arg
 85 90 95
 Asp Arg Val Leu Met Tyr Gly Gln Pro Tyr Arg Val Thr Leu Glu Leu
 100 105 110
 Glu Leu Pro Glu Ser Pro Val Asn Gln Asp Leu Gly Met Phe Leu Val
 115 120 125
 Thr Ile Ser Cys Tyr Thr Arg Gly Gly Arg Ile Ile Ser Thr Ser Ser
 130 135 140
 Arg Ser Val Met Leu His Tyr Arg Ser Asp Leu Leu Gln Met Leu Asp
 145 150 155 160
 Thr Leu Val Phe Ser Ser Leu Leu Leu Phe Gly Phe Ala Glu Gln Lys
 165 170 175
 Gln Leu Leu Glu Val Glu Leu Tyr Ala Asp Tyr Arg Glu Asn Ser Tyr
 180 185 190
 Val Pro Thr Thr Gly Ala Ile Ile Glu Ile His Ser Lys Arg Ile Gln
 195 200 205
 Leu Tyr Gly Ala Tyr Leu Arg Ile His Ala His Phe Thr Gly Leu Arg
 210 215 220
 Tyr Leu Leu Tyr Asn Phe Pro Met Thr Cys Ala Phe Ile Gly Val Ala
 225 230 235 240
 Ser Asn Phe Thr Phe Leu Ser Val Ile Val Leu Phe Ser Tyr Met Gln
 245 250 255
 Trp Val Trp Gly Gly Ile Trp Pro Arg His Arg Phe Ser Leu Gln Val
 260 265 270
 Asn Ile Arg Lys Arg Asp Asn Ser Arg Lys Glu Val Gln Arg Arg Ile
 275 280 285
 Ser Ala His Gln Pro Gly Pro Glu Gly Gln Glu Glu Ser Thr Pro Gln
 290 295 300

Ser Asp Val Thr Glu Asp Gly Glu Ser Pro Glu Asp Pro Ser Gly Thr
 305 310 315 320
 Glu Val Ser Cys Pro Arg Arg Arg Asn Gln Ile Ser Ser Pro
 325 330

(2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 276 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

Met Thr His Pro Gly Thr Gly Asp Ile Ile Ala Val Met Ile Thr Glu
 1 5 10 15
 Leu Arg Gly Lys Asp Ile Leu Ser Tyr Leu Glu Lys Asn Ile Ser Val
 20 25 30
~~Gln Met Thr Ile Ala Val Gly Thr Arg Met Pro Pro Lys Asn Phe Ser~~
 35 40 45
 Arg Gly Ser Leu Val Phe Val Ser Ile Ser Phe Ile Val Leu Met Ile
 50 55 60
 Ile Ser Ser Ala Trp Leu Ile Phe Tyr Phe Ile Gln Lys Ile Arg Tyr
 65 70 75 80
 Thr Asn Ala Arg Asp Arg Asn Gln Arg Arg Leu Gly Asp Ala Ala Lys
 85 90 95
 Lys Ala Ile Ser Lys Leu Thr Thr Arg Thr Val Lys Lys Gly Asp Lys
 100 105 110
 Glu Thr Asp Pro Asp Phe Asp His Cys Ala Val Cys Ile Glu Ser Tyr
 115 120 125
 Lys Gln Asn Asp Val Val Arg Ile Leu Pro Cys Lys His Val Phe His
 130 135 140
 Lys Ser Cys Val Asp Pro Trp Leu Ser Glu His Cys Thr Cys Pro Met

```

145          150          155          160
Cys Lys Leu Asn Ile Leu Lys Ala Leu Gly Ile Val Pro Asn Leu Pro
          165          170          175
Cys Thr Asp Asn Val Ala Phe Asp Met Glu Arg Leu Thr Arg Thr Gln
          180          185          190
Ala Val Asn Arg Arg Ser Ala Leu Gly Asp Leu Ala Gly Asp Asn Ser
          195          200          205
Leu Gly Leu Glu Pro Leu Arg Thr Ser Gly Ile Ser Pro Leu Pro Gln
          210          215          220
Asp Gly Glu Leu Thr Pro Arg Thr Gly Glu Ile Asn Ile Ala Val Thr
225          230          235          240
Lys Glu Trp Phe Ile Ile Ala Ser Phe Gly Leu Leu Ser Ala Leu Thr
          245          250          255
Leu Cys Tyr Met Ile Ile Arg Ala Thr Ala Ser Leu Asn Ala Asn Glu
          260          265          270
Val Glu Trp Phe
          275

```

(2) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 210 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

```

Met Ala Asn Ser Gly Leu Gln Leu Leu Gly Phe Ser Met Ala Leu Leu
 1              5              10              15
Gly Trp Val Gly Leu Val Ala Cys Thr Ala Ile Pro Gln Trp Gln Met
          20              25              30
Ser Ser Tyr Ala Gly Asp Asn Ile Ile Thr Ala Gln Ala Met Tyr Lys
          35              40              45

```

Gly Leu Trp Met Asp Cys Val Thr Gln Ser Thr Gly Met Met Ser Cys
 50 55 60
 Lys Met Tyr Asp Ser Val Leu Ala Leu Ser Ala Ala Leu Gln Ala Thr
 65 70 75 80
 Arg Ala Leu Met Val Val Ser Leu Val Leu Gly Phe Leu Ala Met Phe
 85 90 95
 Val Ala Thr Met Gly Met Lys Cys Thr Arg Cys Gly Gly Asp Asp Lys
 100 105 110
 Val Lys Lys Ala Arg Ile Ala Met Gly Gly Gly Ile Ile Phe Ile Val
 115 120 125
 Ala Gly Leu Ala Ala Leu Val Ala Cys Ser Trp Tyr Gly His Gln Ile
 130 135 140
 Val Thr Asp Phe Tyr Asn Pro Leu Ile Pro Thr Asn Ile Lys Tyr Glu
 145 150 155 160
 Phe Gly Pro Ala Ile Phe Ile Gly Trp Ala Gly Ser Ala Leu Val Ile
 165 170 175
 Leu Gly Gly Ala Leu Leu Ser Cys Ser Cys Pro Gly Asn Glu Ser Lys
 180 185 190
 Ala Gly Tyr Arg Ala Pro Arg Ser Tyr Pro Lys Ser Asn Ser Ser Lys
 195 200 205
 Glu Tyr
 210

(2) INFORMATION FOR SEQ ID NO:37:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 476 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

Met Ile Arg Pro Gln Leu Arg Thr Ala Gly Leu Gly Arg Cys Leu Leu

1	5	10	15
Pro Gly Leu	Leu Leu Leu Val	Pro Val Leu Trp	Ala Gly Ala Glu
20	25	30	
Lys Leu His	Thr Gln Pro Ser Cys	Pro Ala Val Cys	Gln Pro Thr Arg
35	40	45	
Cys Pro Ala	Leu Pro Thr Cys	Ala Leu Gly Thr	Thr Pro Val Phe Asp
50	55	60	
Leu Cys Arg	Cys Cys Arg Val	Cys Pro Ala Ala	Glu Arg Glu Val Cys
65	70	75	80
Gly Gly Ala	Gln Gly Gln Pro	Cys Ala Pro Gly	Leu Gln Cys Leu Gln
85	90	95	
Pro Leu Arg	Pro Gly Phe Pro	Ser Thr Cys Gly	Cys Pro Thr Leu Gly
100	105	110	
Gly Ala Val	Cys Gly Ser Asp	Arg Arg Thr Tyr	Pro Ser Met Cys Ala
115	120	125	
Leu Arg Ala	Glu Asn Arg Ala	Ala Arg Arg Leu	Gly Lys Val Pro Ala
130	135	140	
Val Pro Val	Gln Trp Gly Asn	Cys Gly Asp Thr	Gly Thr Arg Ser Ala
145	150	155	160
Gly Pro Leu	Arg Arg Asn Tyr	Asn Phe Ile Ala	Ala Val Val Glu Lys
165	170	175	
Val Ala Pro	Ser Val Val His	Val Gln Leu Trp	Gly Arg Leu Leu His
180	185	190	
Gly Ser Arg	Leu Val Pro Val	Tyr Ser Gly Ser	Gly Phe Ile Val Ser
195	200	205	
Glu Asp Gly	Leu Ile Ile Thr	Asn Ala His Val	Val Arg Asn Gln Gln
210	215	220	
Trp Ile Glu	Val Val Leu Gln	Asn Gly Ala Arg	Tyr Glu Ala Val Val
225	230	235	240
Lys Asp Ile	Asp Leu Lys Leu	Asp Leu Ala Val	Ile Lys Ile Glu Ser
245	250	255	
Asn Ala Glu	Leu Pro Val Leu	Met Leu Gly Arg	Ser Ser Asp Leu Arg
260	265	270	
Ala Gly Glu	Phe Val Val Ala	Leu Gly Ser Pro	Phe Ser Leu Gln Asn
275	280	285	
Thr Ala Thr	Ala Gly Ile Val	Ser Thr Lys Gln	Arg Gly Gly Lys Glu

290	295	300
Leu Gly Met Lys Asp Ser Asp Met Asp Tyr Val Gln Ile Asp Ala Thr		
305	310	315
Ile Asn Tyr Gly Asn Ser Gly Gly Pro Leu Val Asn Leu Asp Gly Asp		320
	325	330
Val Ile Gly Val Asn Ser Leu Arg Val Thr Asp Gly Ile Ser Phe Ala		335
	340	345
Ile Pro Ser Asp Arg Val Arg Gln Phe Leu Ala Glu Tyr His Glu His		350
	355	360
Gln Met Lys Gly Lys Ala Phe Ser Asn Lys Lys Tyr Leu Gly Leu Gln		365
	370	375
Met Leu Ser Leu Thr Val Pro Leu Ser Glu Glu Leu Lys Met His Tyr		380
	385	390
Pro Asp Phe Pro Asp Val Ser Ser Gly Val Tyr Val Cys Lys Val Val		395
	405	410
Glu Gly Thr Ala Ala Gln Ser Ser Gly Leu Arg Asp His Asp Val Ile		415
	420	425
Val Asn Ile Asn Gly Lys Pro Ile Thr Thr Thr Thr Asp Val Val Lys		430
	435	440
Ala Leu Asp Ser Asp Ser Leu Ser Met Ala Val Leu Arg Gly Lys Asp		445
	450	455
Asn Leu Leu Leu Thr Val Ile Pro Glu Thr Ile Asn		460
465	470	475

(2) INFORMATION FOR SEQ ID NO:38:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 266 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

Met Val Lys Val Thr Phe Asn Ser Ala Leu Ala Gln Lys Glu Ala Lys
 1 5 10 15
 Lys Asp Glu Pro Glu Ser Gly Glu Glu Ala Leu Ile Ile Pro Pro Asp
 20 25 30
 Ala Val Ala Val Asp Cys Lys Asp Pro Asp Asp Val Val Pro Val Gly
 35 40 45
 Gln Arg Arg Ala Trp Cys Trp Cys Met Cys Phe Gly Leu Ala Phe Met
 50 55 60
 Leu Ala Gly Val Ile Leu Gly Gly Ala Tyr Leu Tyr Lys Tyr Phe Ala
 65 70 75 80
 Leu Gln Pro Asp Asp Val Tyr Tyr Cys Gly Ile Lys Tyr Ile Lys Asp
 85 90 95
 Asp Val Ile Leu Asn Glu Pro Ser Ala Asp Ala Pro Ala Ala Leu Tyr
 100 105 110
 Gln Thr Ile Glu Glu Asn Ile Lys Ile Phe Glu Glu Glu Glu Val Glu
 115 120 125
 Phe Ile Ser Val Pro Val Pro Glu Phe Ala Asp Ser Asp Pro Ala Asn
 130 135 140
 Ile Val His Asp Phe Asn Lys Lys Leu Thr Ala Tyr Leu Asp Leu Asn
 145 150 155 160
 Leu Asp Lys Cys Tyr Val Ile Pro Leu Asn Thr Ser Ile Val Met Pro
 165 170 175
 Pro Arg Asn Leu Leu Glu Leu Leu Ile Asn Ile Lys Ala Gly Thr Tyr
 180 185 190
 Leu Pro Gln Ser Tyr Leu Ile His Glu His Met Val Ile Thr Asp Arg
 195 200 205
 Ile Glu Asn Ile Asp His Leu Gly Phe Phe Ile Tyr Arg Leu Cys His
 210 215 220
 Asp Lys Glu Thr Tyr Lys Leu Gln Arg Arg Glu Thr Ile Lys Gly Ile
 225 230 235 240
 Gln Lys Arg Glu Ala Ser Asn Cys Phe Ala Ile Arg His Phe Glu Asn
 245 250 255
 Lys Phe Ala Val Glu Thr Leu Ile Cys Ser
 260 265

We Claim:

1. An isolated and purified human protein having an amino acid sequence selected from the group consisting of the amino acid sequences shown in SEQ ID Nos:20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, and 38.

2. An isolated and purified human protein having an amino acid sequence which is at least 85% identical to an amino acid sequence selected from the group consisting of the amino acid sequences shown in SEQ ID Nos:20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, and 38.

3. An isolated and purified human polypeptide comprising at least 6 contiguous amino acids of an amino acid sequence selected from the group consisting of the amino acid sequences shown in SEQ ID Nos:20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, and 38.

4. A fusion protein comprising a first protein segment and a second protein segment fused together by means of a peptide bond, wherein the first protein segment consists of at least 6 contiguous amino acids selected from the group consisting of the amino acid sequences shown in SEQ ID Nos:20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, and 38.

5. A preparation of antibodies which specifically bind to the human protein of claim 1.

6. An isolated and purified subgenomic polynucleotide having a nucleotide sequence selected from the group consisting of the nucleotide sequences shown in SEQ ID NOs:1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, and 19.

7. An isolated gene corresponding to a cDNA sequence selected from the group consisting of the nucleotide sequences shown in SEQ ID NOs:1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, and 19.

8. A DNA construct for expressing all or a portion of a human protein having an amino acid sequence selected from the group consisting of the amino acid

sequences shown in SEQ ID Nos:20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, and 38, comprising:

a promoter; and

a polynucleotide segment encoding at least 6 contiguous amino acids of the human protein, wherein the polynucleotide segment is located downstream from the promoter, wherein transcription of the polynucleotide segment initiates at or 3' to the promoter.

9. A host cell comprising a DNA construct comprising:

a promoter; and

a polynucleotide segment encoding at least 6 contiguous amino acids of a human protein having an amino acid sequence selected from the group consisting of the amino acid sequences shown in SEQ ID NOs:20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, and 38, wherein the polynucleotide segment is located downstream from the promoter and wherein transcription of the polynucleotide segment initiates at or 3' to the promoter.

10. A homologously recombinant cell having incorporated therein a new transcription initiation unit, wherein the new transcription initiation unit comprises in 5' to 3' order:

(a) an exogenous regulatory sequence;

(b) an exogenous exon; and

(c) a splice donor site,

wherein the transcription initiation unit is located upstream to a coding sequence of a gene, wherein the gene comprises a nucleotide sequence selected from the group consisting of the nucleotide sequences shown in SEQ ID NOs:1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, and 19 and wherein the exogenous regulatory sequence controls transcription of the coding sequence of the gene.

11. A method of producing a human protein, comprising the steps of:

growing a culture of a cell comprising a DNA construct comprising

(1) a promoter and (2) a polynucleotide segment encoding at least 6 contiguous amino acids of a human protein having an amino acid sequence selected from the

group consisting of the amino acid sequences shown in SEQ ID NOs:20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, and 38, wherein the polynucleotide segment is located downstream from the promoter and wherein transcription of the polynucleotide segment initiates at or 3' to the promoter; and
5 purifying the protein from the culture.

12. A method of producing a human protein, comprising the steps of:
growing a culture of a homologously recombinant cell having
incorporated therein a new transcription initiation unit, wherein the new
transcription initiation unit comprises in 5' to 3' order:

- 10 (a) an exogenous regulatory sequence;
(b) an exogenous exon; and
(c) a splice donor site,

wherein the transcription initiation unit is located upstream to a coding sequence of a gene, wherein the gene comprises a nucleotide sequence selected from the group
15 consisting of the nucleotide sequences shown in SEQ ID NOs:1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, and 19 and wherein the exogenous regulatory sequence controls transcription of the coding sequence of the gene; and
purifying the protein from the culture.

13. A method of identifying a secreted polypeptide which is modified by
20 rough microsomes, comprising the steps of:

transcribing *in vitro* a population of cDNA molecules whereby a
population of cRNA molecules is formed;

translating a first portion of the population of cRNA molecules *in vitro*
25 in the absence of rough microsomes whereby a first population of polypeptides is formed;

translating a second portion of the population of cRNA molecules *in vitro* in
the presence of rough microsomes whereby a second population of polypeptides is
formed;

30 comparing the first population of polypeptides with the second
population of polypeptides; and

detecting polypeptide members of the second population which have been modified by the rough microsomes.

THIS PAGE BLANK (USPTO)